

Selectivity in Probabilistic Causality: Drawing Arrows from Inputs to Stochastic Outputs

Ehtibar N. Dzhafarov*
Purdue University

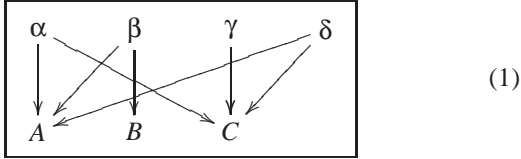
Janne V. Kujala
University of Jyväskylä

Given a set of several inputs into a system (e.g., independent variables characterizing stimuli) and a set of several stochastically non-independent outputs (e.g., random variables describing different aspects of responses), how can one determine, for each of the outputs, which of the inputs it is influenced by? The problem has applications ranging from modeling pairwise comparisons to reconstructing mental processing architectures to conjoint testing. A necessary and sufficient condition for a given pattern of selective influences is provided by the Joint Distribution Criterion, according to which the problem of “what influences what” is equivalent to that of the existence of a joint distribution for a certain set of random variables. For inputs and outputs with finite sets of values this criterion translates into a test of consistency of a certain system of linear equations and inequalities (Linear Feasibility Test) which can be performed by means of linear programming. The Joint Distribution Criterion also leads to a metatheoretical principle for generating a broad class of necessary conditions (tests) for diagrams of selective influences. Among them is the class of distance-type tests based on the observation that certain functionals on jointly distributed random variables satisfy triangle inequality.

KEYWORDS: conjoint testing, external factors, joint distribution, probabilistic causality, mental architectures, metrics on random variables, random outputs, selective influence, stochastic dependence, Thurstonian scaling.

1. INTRODUCTION

This paper presents a general methodology of dealing with *diagrams of selective influences*, like this one:



The Greek letters in this diagram represent *inputs*, or *external factors*, e.g., parameters of stimuli whose values can be chosen at will, or randomly vary but can be observed. The capital Roman letters stand for random outputs characterizing reactions of the system (an observer, a group of observers, a technical device, etc.). The arrows show which factor influences which random output. The factors are treated as *deterministic* entities: even if $\alpha, \beta, \gamma, \delta$ in reality vary randomly (e.g., being randomly generated by a computer program, or being concomitant parameters of observations, such as age of respondents), for the purposes of analyzing selective influences the random outputs A, B, C are always viewed as *conditioned* upon various combinations of specific values of $\alpha, \beta, \gamma, \delta$.

The first question to ask is: what is the meaning of the above diagram if the random outputs A, B, C in it are not necessarily stochastically independent? (If they are, the answer is of course trivial.) And once the meaning of the diagram of selective influences is established, how can one determine that this diagram correctly characterizes the dependence of the joint distributions of the random outputs A, B, C on the external factors

$\alpha, \beta, \gamma, \delta$? These questions are important, because the assumption of stochastic independence of the outputs more often than not is either demonstrably false or adopted for expediency alone, with no other justification, while the assumption of selectivity in causal relations between inputs and stochastic outputs is ubiquitous in theoretical modeling, often being built in the very language of the models.

1.1. An illustration: Pairwise comparisons

Consider Thurstone’s most general model of pairwise comparisons (Thurstone, 1927).¹ This model is predicated on the diagram

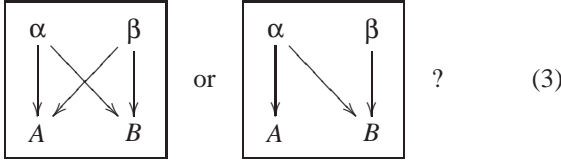


where (A, B) are bivariate normally distributed random variables, and α, β are two stimuli being compared. The stimuli are identified by their “observation areas” (Dzhafarov, 2002): say, the label α may stand for “chronologically first” or “located to the left from fixation point,” and the label β for, respectively, “chronologically second” or “located to the right from fixation point.” For our present purposes, α and β are external factors with varying values (e.g., light intensity in, respectively, first and second observation areas). The random variables A and B

*Corresponding author: Ehtibar Dzhafarov, Purdue University, Department of Psychological Sciences, 703 Third Street West Lafayette, IN 47907, USA. email: ehtibar@purdue.edu. This research has been supported by AFOSR grant FA9550-09-1-0252 to Purdue University and by the Academy of Finland grant 121855 to University of Jyväskylä.

¹ This model is known as Thurstonian Cases 1 and 2. The only difference between the two is that in Case 1 the responding system is an individual observer to whom pairs of stimuli are presented repeatedly, while in Case 2 the responding system is a group of people each responding to every pair of stimuli once. One can, of course, think of all kinds of mixed or intermediate situations.

are supposed to represent some unidimensional property (say, brightness) of the images of, *respectively*, the stimuli α and β (the emphasized word “respectively” indicating selectiveness). According to the model, the probability with which α is judged to have less of the property in question than β equals $\Pr[A < B]$. The problem is: what restrictions should be imposed in this theoretical scheme on the bivariate-normal distribution of A, B to ensure that A is an image of the stimulus α alone and B is an image of the stimulus β alone, as opposed to both or either of them being an image of both the stimuli α and β ? In other words, how can one distinguish, within the framework of Thurstone’s general model, the diagram of selective influences (2) from the diagrams



Denoting by $A(x, y), B(x, y)$ the two random variables at the values (x, y) of the factors (α, β) ,² intuition tells us that one should be able to write

$$A(x, y) = A(x), B(x, y) = B(y)$$

if the diagram (2) holds, but not in the case of the diagrams 3. Clearly then, one should require that

$$\begin{aligned} E[A(x, y)] &= \mu_A(x), \quad \text{Var}[A(x, y)] = \sigma_{AA}(x), \\ E[B(x, y)] &= \mu_B(y), \quad \text{Var}[B(x, y)] = \sigma_{BB}(y), \end{aligned} \quad (4)$$

with the obvious notation for the parameters of the two distributions. These equations form an instance of what is called *marginal selectivity* (the notion introduced in Townsend & Schweickert, 1989) in the dependence of (A, B) on (α, β) : separately taken, the distribution of A (here, normal) does not depend on β , nor the distribution of B on α . The problem is, however, in dealing with the covariance $\text{Cov}[A(x, y), B(x, y)]$. If it is zero for all x, y (i.e., A and B are always stochastically independent), the marginal selectivity is all one needs to speak of α selectively causing A and β selectively causing B . In general, however, the covariance depends on both x and y ,

$$\text{Cov}[A(x, y), B(x, y)] = \sigma_{AB}(x, y).$$

It would be unsatisfactory to simply ignore stochastic interdependence among random variables and focus on marginal selectivity alone. It will be shown in Section 3.3 that marginal selectivity is too weak a concept to allow one to write $A(x, y) = A(x), B(x, y) = B(y)$, because $A(x)$ generally does not preserve its identity (is not the same random variable) under different y ,

and analogously for $B(y)$ under different x . So one needs to answer the conceptual question: under what forms of the dependence of σ_{AB} on (x, y) can one say that the diagram (2) is correct? Even in the seemingly simple special cases one cannot reply on one’s common sense alone. Thus, if $\sigma_{AB}(x, y) = \sigma_{AB}(x)$, what does this tell us about the selectiveness? Even simpler: what can one conclude if one finds out that $\sigma_{AB}(x, y) = \text{const} \neq 0$ across all x, y ? After all, if σ_{AB} is a constant, other measures of stochastic interdependence will be functions of both x and y . For instance, the correlation coefficient then is

$$\text{Cor}[A(x, y), B(x, y)] = \frac{\text{const}}{\sqrt{\sigma_{AA}(x)\sigma_{BB}(y)}} = \rho(x, y).$$

One might be tempted to adopt a radical solution: to always attribute each of A and B to both α and β (i.e., deny any selectiveness), unless A and B are stochastically independent and exhibit marginal selectivity. But a simple example will show that such an approach would be far too restrictive to be useful.

Consider the model in which the observer can be in one of two states of attention, or activation, called “attentive” and “inattentive,” with probabilities p and $1 - p$, respectively. When in the inattentive state, the stimuli α, β (with respective values x, y) cause independent normally distributed images $A(x), B(y)$, with parameters

$$E[A(x)] = 0, \quad \text{Var}[A(x)] = 1,$$

$$E[B(y)] = 0, \quad \text{Var}[B(y)] = 1.$$

That is, in the inattentive state the distribution of the images does not depend on the stimuli at all. When in the attentive state, $A(x), B(y)$ remain independent and normally distributed, but their parameters change as

$$E[A(x)] = \mu_A(x), \quad \text{Var}[A(x)] = 1,$$

$$E[B(y)] = \mu_B(y), \quad \text{Var}[B(y)] = 1.$$

We note that, first, A and B are stochastically independent in either state of attention; second, that A does not depend on β and B does not depend on α in either state of attention; and third, that the switches from one attention state to another do not depend on the stimuli at all. It is intuitively clear then that the causality is selective here, in conformity with the diagram 2. But the overall distribution of A, B in this example (a mixture of two bivariate normal distributions), while obviously satisfying marginal selectivity, has

$$\text{Cov}[A(x, y), B(x, y)] = p(1 - p)\mu_A(x)\mu_B(y) \neq 0.$$

In the theory of selectiveness presented later in this paper it is easily proved that in this situation A only depends on α and B on β , in spite of their stochastic interdependence (see Example 2.5).

It is instructive to see that if one ignores the issue of selectiveness and formulates Thurstone’s general model as Thurstone did it himself, with no restrictions imposed on the covariance $\sigma_{AB}(x, y)$, the model becomes redundant and unfalsifiable, not just with respect to a finite matrix of data, but for any theoretical

² It may seem unnecessary to use separate notation for factors and their values (levels), but it is in fact more convenient in view of the formal treatment presented below. The factors there are defined as sets of “factor points,” and the latter are defined as factor values associated with particular factor names: e.g., (x, α) is a factor point of factor α .

probability function

$$p(x, y) = \Pr[A(x, y) < B(x, y)] \\ = \Phi\left(\frac{\mu_B(y) - \mu_A(x)}{\sqrt{\sigma_{AA}(x) + \sigma_{BB}(y) + 2\sigma_{AB}(x, y)}}\right), \quad (5)$$

where Φ is the standard normal integral. Denoting $z(x, y) = \Phi^{-1}(p(x, y))$, let $\mu_A(x)$ and $\mu_B(y)$ be any functions such that

$$\left|\frac{\mu_A(x) - \mu_B(y)}{z(x, y)}\right| < M,$$

for some M . Then, putting $\sigma_{AA}(x) \equiv \sigma_{BB}(y) \equiv M^2/2$, one can always find the covariance $\sigma_{AB}(x, y)$ to satisfy (5). On a moment's reflection, this is what one should expect: without the assumption of selective influences Thurstone's general model is essentially the same as the vacuous "model" in which stimuli α and β evoke a single normally distributed random variable $D(x, y)$ (interpretable as "subjective difference" between the value x of α and the value y of β), with the decision rule "say that β exceeds α (in a given respect) if $D(x, y) < 0$, otherwise say that α exceeds β ."

The importance of having a principled way of selectively attributing stochastic images to stimuli they represent is even more obvious in the context of the Thurstonian-type models applied to same-different rather than greater-less judgments (Dzhafarov, 2002). When combined with another constraint, called the "well-behavedness" of the random variables representing stimuli, the notion of selective influences has been shown to impose highly non-obvious constraints on the minima of discrimination functions and the relationship " x of α is the best match for y of β " (for details, see Dzhafarov, 2003b-c, 2006; Kujala & Dzhafarov, 2009)

1.2. History and related notions

Historically, the notion of selective probabilistic causality was introduced in psychology by Sternberg (1969), in the context of the reconstruction of "stages" of mental processing. If α and β are certain experimental manipulations (say, size of memory lists and legibility of items, respectively), and if A and B are durations of two hypothetical stages of processing (say, memory search and perception, respectively), then one can hope to test this hypothesis (that memory search and perception are indeed two stages, processes occurring one after another) only if one assumes that A is selectively influenced by α and B by β . Sternberg allows for the possibility of A and B being stochastically interdependent, but it seems that in this case he reduces the selectivity of the influence of α, β upon A, B to a condition that is weaker than even marginal selectivity: the condition is that the mean value of A only depends on α and the mean value of B on β , while any other parameter of the distributions of A and B , say, variance, may very well depend on both α and β .

Townsend (1984), basing his analysis on Townsend and Ashby (1983, Chapter 12), was the first to investigate the notion of selective influences without assuming that the processes which may be selectively influenced by factors are organized serially. He proposed to formalize the notion of selectively influenced and stochastically interdependent random variables by the

concept of "indirect nonselectiveness": the conditional distribution of the variable A given any value b of the variable B , depends on α only, and, by symmetry, the conditional distribution of B at any $A = a$ depends on β only. Under the name of "conditionally selective influence" this notion was mathematically characterized and generalized in Dzhafarov (1999). Although interesting in its own right, this notion turns out to be inadequate, however, for capturing even the most obvious desiderata for the notion of selective influences. In particular, indirect nonselectiveness does not imply marginal selectivity, in fact is not even compatible with it in nontrivial cases. Consider Thurstone's general model again. If both the indirect nonselectiveness and marginal selectivity are satisfied, then

$$E[A|B = b] = \mu_A(x) + \frac{\sigma_{AB}(x, y)}{\sigma_{BB}(y)}(b - \mu_B(y)) = \mu_{A|b}(x),$$

$$\text{Var}[A|B = b] = \left(1 - \frac{\sigma_{AB}^2(x, y)}{\sigma_{AA}(x)\sigma_{BB}(y)}\right)\sigma_{AA}(x) = \sigma_{AA|b}(x),$$

$$E[B|A = a] = \mu_B(y) + \frac{\sigma_{AB}(x, y)}{\sigma_{AA}(x)}(a - \mu_A(x)) = \mu_{B|a}(y),$$

$$\text{Var}[B|A = a] = \left(1 - \frac{\sigma_{AB}^2(x, y)}{\sigma_{AA}(x)\sigma_{BB}(y)}\right)\sigma_{BB}(y) = \sigma_{BB|a}(y).$$

It is not difficult to show that these equations can be satisfied if and only if either

- (i) $\sigma_{AB}(x, y) \equiv 0$, in which case the notions of indirect nonselectiveness and of marginal selectivity simply coincide; or
- (ii) the joint distribution of (A, B) does not depend on either α or β (i.e., $\mu_A, \mu_B, \sigma_{AA}, \sigma_{BB}$, and σ_{AB} are all constants).

Neither of these cases, of course, calls for indirect nonselectiveness as a separate notion.

The difficulty of developing a rigorous and useful definition of selective influences has nothing to do with the fact that in the above examples the random outputs in the diagrams of selective influences are unobservable. They may very well be entirely observable, at least on a sample level. An example would be two performance tests, with outcomes A and B , conducted on a group of people divided into four subgroups according as they were trained or not trained for the A -test and for the B -test. It may be reasonable to hypothesize (at least for some pairs of tests) that the random test score A is selectively influenced by the factor α with the values 'not trained for the A -test' and 'trained for the A -test', while the random test score B is selectively influenced by the factor β with the values 'not trained for the B -test' and 'trained for the B -test'. It is highly likely, however, that the values of A and B will be stochastically interdependent within each of the four subgroups.

A definition of selective influences we adopt in this paper was proposed in Dzhafarov (2003a), and further developed in Dzhafarov and Gluhovsky (2006), Kujala and Dzhafarov (2008), and

Dzhafarov and Kujala (2010). Its rigorous formulation is given in Section 2, but the gist of it, when applied to a diagram like (2), is as follows: there is a random entity R whose distribution does not depend on either of the factors α, β , such that A can be presented as a transformation of R determined by the value x of α , and B can be presented as a transformation of R determined by the value y of β , so that for every allowable pair x, y , the joint distribution of A, B at these x, y is the same as the joint distributions of the two corresponding transformations of R . In the case of the diagram (1), the transformations are

$$f_1(R, x, y, u), f_2(R, y), f_3(R, x, z, u),$$

where x, y, z, u are values of $\alpha, \beta, \gamma, \delta$, respectively.

With some additional assumptions this definition has been applied to Thurstonian-type modeling for same-different comparisons (Dzhafarov, 2003b-c; Kujala & Dzhafarov, 2009), as well as to the hypothetical networks of processes underlying response times (Dzhafarov, Schweickert, Sung, 2004; Schweickert, Fisher, & Goldstein, 2010). Unexplicated, intuitive uses of this notion's special versions can even be found in much earlier publications, such as Bloxom (1972), Schweickert (1982), and Dzhafarov (1992, 1997). In the latter two publications, for instance, response time is considered the sum of a signal-dependent and a signal-independent components, whose durations may very well be stochastically interdependent (even perfectly positively correlated).

Any combination of regression-analytic and factor-analytic models can be viewed as a special version of our definition of selective influences. When applied to the diagram (1), such a model would have the form

$$f_1(R, x, y, u) = h_1(C, x, y, u) + g_1(x, y, u)S_1,$$

$$f_2(R, y) = h_2(C, y) + g_2(y)S_2,$$

$$f_3(R, y, z, u) = h_3(C, y, z, u) + g_3(y, z, u)S_3,$$

where C is a vector of random variables ("common sources of variation"), S_1, S_2, S_3 are "specific sources of variation," all sources of variation being stochastically independent. To recognize in this model our definition one should put $R = (C, S_1, S_2, S_3)$. With some distributional assumptions, this model, for every possible quadruple (x, y, z, u) , has the structure of the nonlinear factor analysis (McDonald, 1967, 1982); the more familiar linear structure is obtained by making h_1, h_2, h_3 linear in the components of C .³

More details on the early history of the notion of selective influences can be found in Dzhafarov (2003a). The relation of this notion to that of "probabilistic explanation" in the sense of Suppes and Zanotti (1982) and to that of "probabilistic dimensionality" in psychometrics (Levine, 2003) are discussed in Dzhafarov

and Gluhovsky (2006). The probabilistic foundations of the issues involved are elaborated in Dzhafarov and Gluhovsky (2006) and, especially, Dzhafarov and Kujala (2010).

Plan of the paper

In this paper we are primarily concerned with necessary (and, under additional constraints, necessary and sufficient) conditions for diagrams of selective influences, like (1) or (2). We call these conditions "tests," in the same way in mathematics we speak of the tests for convergence or for divisibility. That is, the meaning of the term is non-statistical. We assume that random outputs are known on the population level. The principles of constructing statistical tests based on our population level tests are discussed in Section 3.4.2, but specific statistical issues are outside the scope of this paper.

Unlike in Dzhafarov and Kujala (2010), we do not pursue the goal of maximal generality of formulations, focusing instead on the conceptual set-up that would apply to commonly encountered experimental designs. This means a finite number of factors, each having a finite number of values, with some (not necessarily all) combinations of the values of the factors serving as allowable treatments. It also means that the random outcomes influenced by these factors are *random variables*: their values are vectors of real numbers or elements of countable sets, rather than more complex structures, such as functions or sets. To keep the paper self-contained, however, we have added an appendix in which we formulate the main definitions and statements of the theory on a much higher level of generality: for arbitrary sets of factors, arbitrary sets of factors values, and arbitrarily complex random outcomes.

In Section 2 we introduce the notion of several random variables influenced by several factors and formulate a definition of selective influences. In Section 3 we present the Joint Distribution Criterion, a necessary and sufficient condition for selective influences (or, if one prefers, an alternative definition thereof), and we list three basic properties of selective influences. In the same section we formulate the principle by which one can construct tests for selective influences, on population and sample levels. In Section 4 we describe the main and universally applicable test for selective influences, Linear Feasibility Test. The test is universally applicable because every random outcome and every set of factors can be discretized into a finite number of categories. The Linear Feasibility Test is both necessary and sufficient condition for selective influences within the framework of the chosen discretization of inputs and outputs. In Section 5 we study tests based on "pseudo-quasi-metrics" defined on spaces of jointly distributed random variables, and we introduce many examples of such tests. Finally, in Section 6 we discuss, with less elaboration, two examples of non-distance-type tests.

2. BASIC NOTIONS

2.1. Factors, factor points, treatments

A *factor* α , formally, is a set of *factor points*, each of which has the format "value (or level) x of factor α ." In symbols, this

³ To avoid confusion, our use of the term "factor" is reserved for observable external inputs (corresponding to the use of the term in MANOVA); the unobservable "factors" of the factor analysis can be referred to in the present context as "sources of variation," or "sources of randomness."

can be presented as (x, α') , where α' is the unique name of the set α rather than the set itself. It is convenient to write x^α in place of (x, α') . Thus, if a factor with the name ‘*intensity*’ has three levels, ‘*low*,’ ‘*medium*,’ and ‘*high*,’ then this factor is taken to be the set

$$\text{intensity} = \{ \text{low}^{\text{intensity}}, \text{medium}^{\text{intensity}}, \text{high}^{\text{intensity}} \}.$$

There is no circularity here, for, say, the factor point $\text{low}^{\text{intensity}}$ stands for $(\text{value} = \text{low}, \text{name} = \text{‘intensity’})$ rather than $(\text{value} = \text{low}, \text{set} = \text{intensity})$.

In the main text we will deal with finite sets of factors $\Phi = \{\alpha_1, \dots, \alpha_m\}$, with each factor $\alpha \in \Phi$ consisting of a finite number of factor points,

$$\alpha = \{v_1^\alpha, \dots, v_{k_\alpha}^\alpha\}.$$

Clearly, $\alpha \cap \beta = \emptyset$ for any distinct $\alpha, \beta \in \Phi$.

A *treatment*, as usual, is defined as the set of factor points containing one factor point from each factor,⁴

$$\phi = \{x_1^{\alpha_1}, \dots, x_m^{\alpha_m}\} \in \alpha_1 \times \dots \times \alpha_m.$$

The *set of treatments* (used in an experiment or considered in a theory) is denoted by $T \subset \alpha_1 \times \dots \times \alpha_m$ and assumed to be nonempty. Note that T need not include all possible combinations of factor points. This is an important consideration in view of the “canonical rearrangement” described below. Also, incompletely crossed designs occur broadly — in an experiment because the entire set $\alpha_1 \times \dots \times \alpha_m$ may be too large, or in a theory because certain combinations of factor points may be physically or logically impossible (e.g., contrast and shape cannot be completely crossed if zero is one of the values for contrast).

Example 2.1. In the diagram (1), let α, β, γ , and δ have respectively 3, 2, 1, and 2 values. Then these factors can be presented as

$$\Phi = \left\{ \begin{array}{l} \alpha = \{1^\alpha, 2^\alpha, 3^\alpha\}, \\ \beta = \{1^\beta, 2^\beta\}, \\ \gamma = \{1^\gamma\}, \\ \delta = \{1^\delta, 2^\delta\} \end{array} \right\}.$$

The only constraint on one’s choice of the labels for the values (here, 1, 2, 3) is that within a factor they should be pairwise distinct. Due to the unique superscripting, no two factors can share a factor point. The maximum number of possible treatments in this example is 12, in which case

$$T = \left\{ \begin{array}{l} \{1^\alpha, 1^\beta, 1^\delta\}, \{1^\alpha, 1^\beta, 2^\delta\}, \{1^\alpha, 2^\beta, 1^\delta\}, \{1^\alpha, 2^\beta, 2^\delta\}, \\ \{2^\alpha, 1^\beta, 1^\delta\}, \{2^\alpha, 1^\beta, 2^\delta\}, \{2^\alpha, 2^\beta, 1^\delta\}, \{2^\alpha, 2^\beta, 2^\delta\}, \\ \{3^\alpha, 1^\beta, 1^\delta\}, \{3^\alpha, 1^\beta, 2^\delta\}, \{3^\alpha, 2^\beta, 1^\delta\}, \{3^\alpha, 2^\beta, 2^\delta\} \end{array} \right\}.$$

⁴ We present treatments as sets $\{x_1^{\alpha_1}, \dots, x_m^{\alpha_m}\}$ rather than vectors $(x_1^{\alpha_1}, \dots, x_m^{\alpha_m})$, which would be a correct representation of elements of $\alpha_1 \times \dots \times \alpha_m$, because the superscripting we use makes the ordering of the points $x_i^{\alpha_i}$ irrelevant.

We have deleted 1^γ from all treatments because a factor with a single factor point can always be removed from a diagram (or added to a diagram, if convenient; see \emptyset^α notation in Section 3.1). \square

2.2. Random variables

A rigorous definition of a *random variable* (as a special case of a random entity) is given in the appendix. For simplicity of notation, any *random variable* A considered in the main text may be assumed to be a vector of “more elementary” *discrete* and *continuous* random variables: for a discrete variable, the set of its possible values is countable (finite or infinite), and each value possesses a *probability mass*; in the continuous case, the set of possible values is \mathbb{R}^N (vectors with N real-valued components), and each $a \in \mathcal{A}$ possesses a conventional *probability density*. So a random variable A consists of several jointly distributed components, (A_1, \dots, A_k) , some (or all) of which are continuous and some (or all) of which are discrete. Note that random vectors in this terminology are random variables. The set of possible values of A is denoted \mathcal{A} and each $a \in \mathcal{A}$ has a mass/density value $p(a)$ associated with it.⁵

Every vector of jointly distributed random variables $A = (A_1, \dots, A_n)$ is a random variable, and every value $a = (a_1, \dots, a_n) \in \mathcal{A}_1 \times \dots \times \mathcal{A}_n$ of this random variable possesses a *joint mass/density* $p(a) = p(a_1, \dots, a_n)$; then for any subvector $(a_{i_1}, \dots, a_{i_k})$ of (a_1, \dots, a_n) the mass/density $p_{i_1, \dots, i_k}(a_{i_1}, \dots, a_{i_k})$ is obtained by summing and/or integrating $p(a_1, \dots, a_n)$ across all possible values of $(a_1, \dots, a_n) - (a_{i_1}, \dots, a_{i_k})$. Note, however, that a vector of random variables $A = (A_1, \dots, A_n)$ need not be a random variable, because (A_1, \dots, A_n) need not possess a joint distribution.

We use the relational symbol \sim in the meaning of “is distributed as.” $A \sim B$ is well defined irrespective of whether A and B are jointly distributed.

Let, for each treatment $\phi \in T$, there be a vector of jointly distributed random variables with the set of possible values $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$ (that does not depend on ϕ) and probability mass/density $p_\phi(a_1, \dots, a_n)$ that depends on ϕ .⁶ Then we say that we have a *vector of jointly distributed random variables that depends on treatment* ϕ , and write

$$A(\phi) = (A_1, \dots, A_n)(\phi), \quad \phi \in T.$$

A correct way of thinking of $A(\phi)$ is that it represents a *set of vectors of jointly distributed random variables*, each of these vectors being labeled (indexed) by a particular treatment. Any

⁵ Probability mass/density is generally the Radon-Nikodym derivative with respect to the product of a counting measure and the Lebesgue measure on \mathbb{R}^N .

⁶ The invariance of \mathcal{A} with respect to ϕ (more generally, the invariance of the observation space for A with respect to ϕ) is convenient to assume, but it is not essential for the theory. Its two justifications are that (a) this requirement makes it natural to speak of “one and the same” A whose distribution changes with ϕ rather than to speak (more correctly) of different random variables $A(\phi)$ for different ϕ ; and (b) in the context of selective influences one can always redefine the observation spaces for different treatments ϕ to make them coincide (see Remark A.6 in the appendix).

subvector of $A(\phi)$ should also be written with the argument ϕ , say, $(A_1, A_2, A_3)(\phi)$. If ϕ is explicated as $\phi = \{x_1^{\alpha_1}, \dots, x_m^{\alpha_m}\}$ or, say, $\phi = \{3^\alpha, 1^\beta, 1^\delta\}$, we will write $A(x_1^{\alpha_1}, \dots, x_m^{\alpha_m})$ or $(A, B, C)(3^\alpha, 1^\beta, 1^\delta)$ instead of more correct $A(\{x_1^{\alpha_1}, \dots, x_m^{\alpha_m}\})$ or $(A, B, C)(\{3^\alpha, 1^\beta, 1^\delta\})$.

It is important to note that for distinct treatments ϕ_1 and ϕ_2 the corresponding $A(\phi_1)$ and $A(\phi_2)$ *do not possess a joint distribution*, they are *stochastically unrelated*. This is easy to understand: since ϕ_1 and ϕ_2 are mutually exclusive conditions for observing values of A , there is no non-arbitrary way of choosing which value $a = (a_1, \dots, a_n)$ observed at ϕ_1 should be paired with which value $a' = (a'_1, \dots, a'_n)$ observed at ϕ_2 . To consider $A(\phi_1)$ and $A(\phi_2)$ stochastically independent and to pair every possible value of $A(\phi_1)$ with every possible value $A(\phi_2)$ is as arbitrary as, say, to consider them positively correlated and to pair every quantile of $A(\phi_1)$ with the corresponding quantile of $A(\phi_2)$.

Example 2.2. In diagram (1), let Φ and T be as in Example 2.1, and let A, B, C be binary, 0/1, variables. Then $(A, B, C)(\phi)$ is defined, for each $\phi = \{x^\alpha, y^\beta, z^\delta\}$, by a table of the following form:

α	β	δ	A	B	C	Pr
x	y	z	0	0	0	p_{000}
			0	0	1	p_{001}
			0	1	0	p_{010}
			0	1	1	p_{011}
			1	0	0	p_{100}
			1	0	1	p_{101}
			1	1	0	p_{110}
			1	1	1	p_{111}

separately for each of the 12 treatments. \square

2.3. Selective influences

Given a set of factors $\Phi = \{\alpha_1, \dots, \alpha_m\}$ and a vector $A(\phi) = (A_1, \dots, A_n)(\phi)$ of random variables depending on treatment, a *diagram of selective influences* is a mapping

$$M : \{1, \dots, n\} \rightarrow 2^\Phi \quad (6)$$

(2^Φ being the set of subsets of Φ), with the interpretation that

$$\Phi_i = M(i)$$

is the subset of factors (which may be empty) *selectively influencing* A_i ($i = 1, \dots, n$). The definition of selective influences is yet to be given (Definition 2.4), but for the moment think simply of arrows drawn from factors to random variables (or vice versa). The subset of factors Φ_i influencing A_i determines, for any treatment $\phi \in T$, the subtreatments ϕ_{Φ_i} defined as

$$\phi_{\Phi_i} = \{x^\alpha \in \phi : \alpha \in \Phi_i\}, \quad i = 1, \dots, n.$$

Subtreatments ϕ_{Φ_i} across all $\phi \in T$ can be viewed as *admissible values* of the subset of factors Φ_i ($i = 1, \dots, n$). Note that ϕ_{Φ_i} is empty whenever Φ_i is empty.

Example 2.3. In the diagram 1, having enumerated A, B, C by 1, 2, 3, respectively, $\Phi_1 = \{\alpha, \beta, \delta\}$, $\Phi_2 = \{\beta\}$, $\Phi_3 = \{\alpha, \gamma, \delta\}$. If the factor points are as in Examples 2.1 and 2.2, then, choosing $\phi = \{3^\alpha, 1^\beta, 1^\gamma, 2^\delta\}$, we have $\phi_{\Phi_1} = \{3^\alpha, 1^\beta, 2^\delta\}$, $\phi_{\Phi_2} = \{1^\beta\}$, and $\phi_{\Phi_3} = \{3^\alpha, 1^\gamma, 2^\delta\}$ (where γ and its only point 1^γ can be omitted everywhere, making, in particular, the treatments ϕ_{Φ_1} and ϕ coincide). \square

The definition below is a special case of the definition of selective influences given in the appendix. This definition will be easier to justify in terms of the Joint Distribution Criterion formulated in the next section.

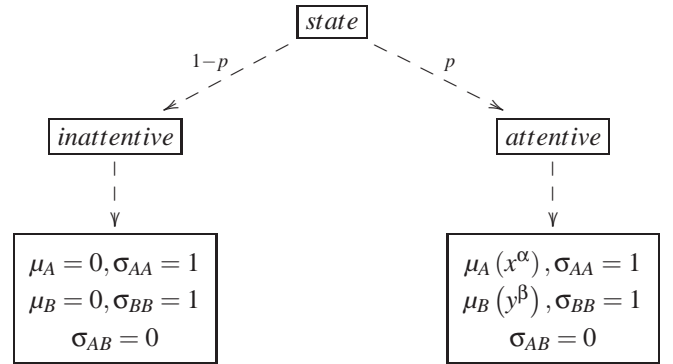
Definition 2.4 (Selective influences). A vector of random variables $A(\phi) = (A_1, \dots, A_n)(\phi)$ is said to satisfy a diagram of selective influences (6) if there is a random variable⁷ R taking values on some set \mathcal{R} , and functions $f_i : \Phi_i \times \mathcal{R} \rightarrow \mathcal{A}_i$ ($i = 1, \dots, n$), such that, for any treatment $\phi \in T$,

$$(A_1, \dots, A_n)(\phi) \sim (f_1(\phi_{\Phi_1}, R), \dots, f_n(\phi_{\Phi_n}, R)). \quad (7)$$

We write then, schematically, $(A_1, \dots, A_n) \leftarrow \Phi (\Phi_1, \dots, \Phi_n)$.

The qualifier “schematically” in reference to $(A_1, \dots, A_n) \leftarrow \Phi (\Phi_1, \dots, \Phi_n)$ is due to the fact that (A_1, \dots, A_n) is not well-defined without mentioning a treatment ϕ at which these variables are taken. This notation, therefore, is merely a compact way of referring to the diagram (6).

Example 2.5. Consider the Thurstonian “mixture” model described in the introduction:



The selectivity $(A, B) \leftarrow (\alpha, \beta)$ here is shown by

⁷ Even though $A(\phi)$ is a random variable, and Φ is a finite set of factors containing a finite set of factor points each, the requirement in the definition that R be a random variable is unnecessarily restrictive: it is sufficient to require the existence of a *random entity* R distributed on some probability space $(\mathcal{R}, \Sigma_{\mathcal{R}}, \mu)$ (see the appendix). It is shown in the appendix, however, based on the Joint Distribution Criterion, that if the definition is satisfied with an arbitrary R , then the latter can always be chosen to be a random variable — discrete, continuous, or mixed according as the variable $A(\phi)$ is discrete, continuous, or mixed. (Recall that in our terminology every vector of random variables is a random variable.) Moreover, R can always be chosen to be distributed unit-uniformly, or according to any distribution function strictly increasing on any interval of reals constituting \mathcal{R} .

1. putting $R = (S, N_1, N_2)$, where S is a Bernoulli (0/1) variable with $\Pr[S = 1] = p$, N_1, N_2 are standard normal variables, and the three variables are independent;
2. defining

$$(f_1(x^\alpha, (S, N_1, N_2)), f_2(y^\beta, (S, N_1, N_2))) \\ = (\mu_A(x^\alpha)S + N_1, \mu_B(y^\beta)S + N_2);$$

3. and observing that

$$(\mu_A(x^\alpha)S + N_1, \mu_B(y^\beta)S + N_2) \sim (A, B)(x^\alpha, y^\beta)$$

for all treatments $\{x^\alpha, y^\beta\}$. \square

Remark 2.6. Note that the components of $(f_1(\phi_{\Phi_1}, R), \dots, f_n(\phi_{\Phi_n}, R))$ are jointly distributed for any given ϕ because they are functions of one and the same random variable. The components of $(A_1, \dots, A_n)(\phi)$ are jointly distributed for any given ϕ by definition. There is, however, no joint distribution of these two vectors, $(f_1(\phi_{\Phi_1}, R), \dots, f_n(\phi_{\Phi_n}, R))$ and $(A_1, \dots, A_n)(\phi)$, for any ϕ ; and, as emphasized earlier, no joint distribution for $(A_1, \dots, A_n)(\phi_1)$ and $(A_1, \dots, A_n)(\phi_2)$, for distinct ϕ_1 and ϕ_2 .

3. JOINT DISTRIBUTION CRITERION

3.1. Canonical Rearrangement

The simplest diagram of selective influences is *bijection*,

$$\begin{array}{ccc} \alpha_1 & \dots & \alpha_n \\ \downarrow & & \downarrow \\ A_1 & \dots & A_n \end{array} \quad (8)$$

In this case we write $(A_1, \dots, A_n) \leftarrow \leftarrow (\alpha_1, \dots, \alpha_n)$ instead of $(A_1, \dots, A_n) \leftarrow (\Phi_1 = \{\alpha_1\}, \dots, \Phi_n = \{\alpha_n\})$.

We can simplify the subsequent discussion without sacrificing generality by agreeing to reduce each diagram of selective influences to a bijective form, by appropriately redefining factors and treatments. It is almost obvious how this should be done. Given the subsets of factors $\Phi_1 \dots, \Phi_n$ determined by a diagram of selective influences (6), each Φ_i can be viewed as a factor identified with the set of factor points

$$\alpha_i^* = \{(\phi_{\Phi_i})^{\alpha_i^*} : \phi \in T\},$$

in accordance with the notation we have adopted for factor points: $(\phi_{\Phi_i})^{\alpha_i^*} = (\phi_{\Phi_i}, \alpha_i^*)$. If Φ_i is empty, then ϕ_{Φ_i} is empty too, and we should designate a certain value, say $\emptyset^{\alpha_i^*}$, as a dummy factor point (the only element of factor α_i^*). The set of treatments T for the original factors $\{\alpha_1, \dots, \alpha_n\}$ should then be redefined for the vector of new factors $(\alpha_1^*, \dots, \alpha_n^*)$ as

$$T^* = \left\{ \left\{ (\phi_{\Phi_1})^{\alpha_1^*}, \dots, (\phi_{\Phi_n})^{\alpha_n^*} \right\} : \phi \in T \right\} \subset \alpha_1^* \times \dots \times \alpha_n^*.$$

We call this redefinition of factor points, factors, and treatments the *canonical rearrangement*.

Example 3.1. Diagram (1), with the factors defined as in Example 2.1 (with γ omitted), is reduced to a bijective form as follows:

$$\alpha^* = \left\{ \left\{ x_1^\alpha, x_2^\beta, x_3^\delta \right\}^{\alpha^*} : \left\{ x_1^\alpha, x_2^\beta, x_3^\delta \right\} \in \alpha \times \beta \times \delta \right\},$$

$$\beta_2^* = \left\{ \left\{ y^\beta \right\}^{\beta^*} : y^\beta \in \beta \right\},$$

$$\gamma_3^* = \left\{ \left\{ z_1^\alpha, z_3^\delta \right\}^{\gamma^*} : \left\{ z_1^\alpha, z_3^\delta \right\} \in \alpha \times \delta \right\},$$

with, respectively, 12, 2, and 6 factor points, and

$$T^* = \left\{ \left\{ \left\{ x_1^\alpha, x_2^\beta, x_3^\delta \right\}^{\alpha^*}, \left\{ y^\beta \right\}^{\beta^*}, \left\{ z_1^\alpha, z_3^\delta \right\}^{\gamma^*} \right\} \in \alpha_1^* \times \beta_2^* \times \gamma_3^* \right\}, \\ : x_1^\alpha = z_1^\alpha, x_2^\beta = y^\beta, x_3^\delta = z_3^\delta \right\},$$

the number of treatments, obviously remaining the same, 12, as for the original factors. \square

The purpose of canonical rearrangement is to achieve a bijective correspondence between factors and the random variables selectively influenced by these factors. Equivalently, we may say that the random variables following canonical rearrangement can be indexed by the factors (assumed to be) selectively influencing them. Thus, if we test the hypothesis that $(A_1, \dots, A_n) \leftarrow (\alpha_1, \dots, \alpha_n)$, we can, when convenient, write $A_{\{\alpha_1\}}$ in place of A_1 , $A_{\{\alpha_2\}}$ in place of A_2 , etc.

3.2. The criterion

From now on let us assume that we deal with bijective diagrams of selective influences, (8). The notation $\phi_{\Phi_i} = \phi_{\{\alpha_i\}}$ then indicates the singleton set $\{x^{\alpha_i}\} \subset \phi$. As usual, we write x^{α_i} in place of $\{x^{\alpha_i}\}$:

$$\phi_{\{\alpha_i\}} = \{x_1^{\alpha_1}, \dots, x_n^{\alpha_n}\}_{\{\alpha_i\}} = x_i^{\alpha_i}.$$

The definition of selective influences (Definition 2.4) then acquires the following form:

Definition 3.2 (Selective influences, bijective form). A vector of random variables $A(\phi) = (A_1, \dots, A_n)(\phi)$ is said to satisfy a diagram of selective influences (8), and we write $(A_1, \dots, A_n) \leftarrow (\alpha_1, \dots, \alpha_n)$, if, for some random variable⁸ R and for any treatment $\phi \in T$,

$$(A_1, \dots, A_n)(\phi) \sim (f_1(\phi_{\{\alpha_1\}}, R), \dots, f_n(\phi_{\{\alpha_n\}}, R)), \quad (9)$$

where $f_i : \alpha_i \times \mathcal{R} \rightarrow \mathcal{A}_i$ ($i = 1, \dots, n$) are some functions, with \mathcal{R} denoting the set of possible values of R .

This definition is difficult to put to work, as it refers to an existence of a random variable R without showing how one can

⁸ See footnote 7.

find it or prove that it cannot be found. In Dzhafarov and Kujala (2010), however, we have formulated a necessary and sufficient condition for $(A_1, \dots, A_n) \leftarrow \mathcal{P} (\alpha_1, \dots, \alpha_n)$ which circumvents this problem.

Criterion 3.3 (Joint Distribution Criterion, JDC). *A vector of random variables $A(\phi) = (A_1, \dots, A_n)(\phi)$ satisfies a diagram of selective influences (8) if and only if there is a vector of jointly distributed random variables*

$$H = \left(\overbrace{H_{x_1^{\alpha_1}}, \dots, H_{x_{k_1}^{\alpha_1}}}^{\text{for } \alpha_1}, \dots, \overbrace{H_{x_1^{\alpha_n}}, \dots, H_{x_{k_n}^{\alpha_n}}}^{\text{for } \alpha_n} \right),$$

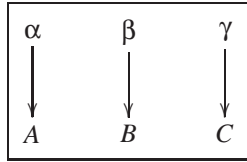
one random variable for each factor point of each factor, such that

$$(H_{\phi_{\{\alpha_1\}}}, \dots, H_{\phi_{\{\alpha_n\}}}) \sim A(\phi) \quad (10)$$

for every treatment $\phi \in T$.

Due to its central role, the simple proof of this criterion (for the general case of arbitrary factors and sets of random entities) is reproduced in the appendix. The vector H in the formulation of the JDC is referred to as the *JDC-vector* for $A(\phi)$, or the *hypothetical JDC-vector* for $A(\phi)$, if the existence of such a vector of jointly distributed variables is in question.

Example 3.4. For the diagram of selective influences



with $\alpha = \{1^\alpha, 2^\alpha\}$, $\beta = \{1^\beta, 2^\beta, 3^\beta\}$, $\gamma = \{1^\gamma, 2^\gamma, 3^\gamma, 4^\gamma\}$, and the set of allowable treatments

$$T = \left\{ \begin{array}{l} \{1^\alpha, 2^\beta, 1^\gamma\}, \{1^\alpha, 2^\beta, 3^\gamma\}, \{2^\alpha, 1^\beta, 4^\gamma\}, \\ \{1^\alpha, 3^\beta, 1^\gamma\}, \{2^\alpha, 3^\beta, 2^\gamma\} \end{array} \right\},$$

the hypothetical JDC-vector is

$$(H_{1^\alpha}, H_{2^\alpha}, H_{1^\beta}, H_{2^\beta}, H_{3^\beta}, H_{1^\gamma}, H_{2^\gamma}, H_{3^\gamma}, H_{4^\gamma}),$$

the hypothesis being that

$$(H_{1^\alpha}, H_{2^\beta}, H_{1^\gamma}) \sim (A, B, C) (1^\alpha, 2^\beta, 1^\gamma),$$

$$(H_{1^\alpha}, H_{2^\beta}, H_{3^\gamma}) \sim (A, B, C) (1^\alpha, 2^\beta, 3^\gamma),$$

$$(H_{2^\alpha}, H_{1^\beta}, H_{4^\gamma}) \sim (A, B, C) (2^\alpha, 1^\beta, 4^\gamma),$$

$$(H_{1^\alpha}, H_{3^\beta}, H_{1^\gamma}) \sim (A, B, C) (1^\alpha, 3^\beta, 1^\gamma),$$

$$(H_{2^\alpha}, H_{3^\beta}, H_{2^\gamma}) \sim (A, B, C) (2^\alpha, 3^\beta, 2^\gamma).$$

This means, in particular, that H_{1^α} and H_{2^α} have the same set of values as A (which, by our convention, does not depend on treatment), the set of values for H_{1^β} , H_{2^β} , and H_{3^β} is the same as that of B , and the set of values for H_{1^γ} , H_{2^γ} , H_{3^γ} , and H_{4^γ} is the same as that of C . \square

The JDC prompts a simple justification for our definition of selective influences. Let $(A, B, C) \leftarrow \mathcal{P} (\alpha, \beta, \gamma)$, as in the previous example, with each factor containing two factor points. Consider all treatments ϕ in which the factor point of α is fixed, say, at 1^α . If $(A, B, C) \leftarrow \mathcal{P} (\alpha, \beta, \gamma)$, then in the vectors of random variables

$$(A, B, C) (1^\alpha, 2^\beta, 1^\gamma), (A, B, C) (1^\alpha, 2^\beta, 3^\gamma), (A, B, C) (1^\alpha, 3^\beta, 1^\gamma),$$

the marginal distribution of the variable A is one and the same,

$$A (1^\alpha, 2^\beta, 1^\gamma) \sim A (1^\alpha, 2^\beta, 3^\gamma) \sim A (1^\alpha, 3^\beta, 1^\gamma).$$

But the intuition of selective influences requires more: that we can denote this variable $A (1^\alpha)$ because it *preserves its identity* (and not just its distribution) no matter what other variables it is paired with, $(B, C) (2^\beta, 1^\gamma)$, $(B, C) (2^\beta, 3^\gamma)$, or $(B, C) (3^\beta, 1^\gamma)$. Analogous statements hold for $A (2^\alpha)$, $B (2^\beta)$, $B (3^\beta)$, $C (1^\gamma)$. The JDC formalizes the intuitive notion of variables “preserving their identity” when entering in various combinations with each other: there are jointly distributed random variables

$$H_{1^\alpha}, H_{2^\alpha}, H_{1^\beta}, H_{2^\beta}, H_{3^\beta}, H_{1^\gamma}, H_{2^\gamma}, H_{3^\gamma}, H_{4^\gamma}$$

whose identity is defined by this joint distribution; when H_{1^α} is combined with random variables H_{2^β} and H_{3^γ} , it forms the triad $(H_{1^\alpha}, H_{2^\beta}, H_{3^\gamma})$ whose distribution is the same as that of $(A, B, C) (1^\alpha, 2^\beta, 1^\gamma)$; when the same random variable H_{1^α} is combined with random variables H_{2^β} and H_{3^γ} , the triad $(H_{1^\alpha}, H_{2^\beta}, H_{3^\gamma})$ is distributed as $(A, B, C) (1^\alpha, 2^\beta, 3^\gamma)$; and so on — the key concept being that it is *one and the same* H_{1^α} which is being paired with other variables, as opposed to different random variables $A (1^\alpha, 2^\beta, 1^\gamma)$, $A (1^\alpha, 2^\beta, 3^\gamma)$, $A (1^\alpha, 3^\beta, 1^\gamma)$ which are identically distributed (cf. Example 3.7 below, which shows that the identity is not generally preserved if all we know is marginal selectivity).

3.3. Three basic properties of selective influences

The three properties in question are immediate consequences of JDC.

3.3.1. Property 1: Nestedness.

For any subset $\{i_1, \dots, i_k\}$ of $\{1, \dots, n\}$, if $(A_1, \dots, A_n) \leftarrow \mathcal{P} (\alpha_1, \dots, \alpha_n)$ then $(A_{i_1}, \dots, A_{i_k}) \leftarrow \mathcal{P} (\alpha_{i_1}, \dots, \alpha_{i_k})$.

Example 3.5. In Example 3.4, if $(A, B, C) \leftarrow \mathcal{P} (\alpha, \beta, \gamma)$, then $(A, C) \leftarrow \mathcal{P} (\alpha, \gamma)$, because the JDC criterion for $(A, B, C) \leftarrow \mathcal{P} (\alpha, \beta, \gamma)$

(α, β, γ) implies that $(H_1^\alpha, H_2^\alpha, H_1^\gamma, H_2^\gamma, H_3^\gamma, H_4^\gamma)$ are jointly distributed, and that

$$\begin{aligned} (H_1^\alpha, H_1^\gamma) &\sim (A, C) (1^\alpha, 1^\gamma), \\ (H_1, H_3^\gamma) &\sim (A, C) (1^\alpha, 3^\gamma), \\ (H_2^\alpha, H_2^\gamma) &\sim (A, C) (2^\alpha, 2^\gamma), \\ (H_2^\alpha, H_4^\gamma) &\sim (A, C) (2^\alpha, 4^\gamma). \end{aligned}$$

Analogously, $(A, B) \Leftarrow (\alpha, \beta)$ and $(B, C) \Leftarrow (\beta, \gamma)$. Statements with \Leftarrow involving a single variable merely indicate the dependence of its distribution on the corresponding factor: thus, $A \Leftarrow \alpha$ simply mean that the distribution of $A(x^\alpha, y^\beta, z^\gamma)$ does not depend on y^β, z^γ . \square

3.3.2. Property 2: Complete Marginal Selectivity

For any subset $\{i_1, \dots, i_k\}$ of $\{1, \dots, n\}$, if $(A_1, \dots, A_n) \Leftarrow (\alpha_1, \dots, \alpha_n)$ then the k -marginal distribution⁹ of $(A_{i_1}, \dots, A_{i_k})(\phi)$ does not depend on points of the factors outside $(\alpha_{i_1}, \dots, \alpha_{i_k})$. In particular, the distribution of A_i only depends on points of α_i , $i = 1, \dots, n$.

This is, of course, a trivial consequence of the nestedness property, but its importance lies in that it provides the easiest to check necessary condition for selective influences.

Example 3.6. Let the factors, factor points, and the set of treatments be as in Example 3.4. Let the distributions of (A, B, C) at the five different treatments be as shown:

α	β	γ	A	B	C	Pr
1	2	1	0	0	0	.2
			0	0	1	.1
			0	1	0	.1
			0	1	1	.1
			1	0	0	.1
			1	0	1	.1
			1	1	0	.1
			1	1	1	.2

α	β	γ	A	B	C	Pr
1	2	3	0	0	0	0
			0	0	1	.3
			0	1	0	.2
			0	1	1	0
			1	0	0	.1
			1	0	1	.1
			1	1	0	.1
			1	1	1	.2

α	β	γ	A	B	C	Pr
2	1	4	0	0	0	.3
			0	0	1	0
			0	1	0	.3
			0	1	1	0
			1	0	0	.3
			1	0	1	0
			1	1	0	0
			1	1	1	.1

α	β	γ	A	B	C	Pr
1	3	1	0	0	0	.4
			0	0	1	.1
			0	1	0	0
			0	1	1	0
			1	0	0	0
			1	0	1	.2
			1	1	0	.1
			1	1	1	.2

α	β	γ	A	B	C	Pr
2	3	2	0	0	0	.2
			0	0	1	.1
			0	1	0	.2
			0	1	1	.1
			1	0	0	.3
			1	0	1	.1
			1	1	0	0
			1	1	1	0

One can check that marginal selectivity holds for all 1-marginals: thus, irrespective of other factor points,

α	A	Pr
1	0	.5
	1	.5

α	A	Pr
2	0	.6
	1	.4

β	B	Pr
1	0	.6
	1	.4

β	B	Pr
2	0	.5
	1	.5

β	B	Pr
3	0	.7
	1	.3

γ	A	Pr
1	0	.5
	1	.5

γ	A	Pr
2	0	.7
	1	.3

γ	A	Pr
3	0	.4
	1	.6

γ	A	Pr
4	0	.9
	1	.1

One can also check that irrespective of the factor point of γ , the 2-marginal (A, B) only depends on α and β :

α	β	A	B	Pr
1	2	0	0	.3
		0	1	.2
		1	0	.2
		1	1	.3

α	β	A	B	Pr
2	1	0	0	.3
		0	1	.3
		1	0	.3
		1	1	.1

α	β	A	B	Pr
1	3	0	0	.5
		0	1	0
		1	0	.2
		1	1	.3

α	β	A	B	Pr
2	3	0	0	.3
		0	1	.3
		1	0	.4
		1	1	0

Marginal selectivity, however, is violated for the 2-marginal (A, C) : if the factor point of β is 2^β ,

α	γ	A	C	Pr
1	1	0	0	.3
		1	0	.2
		0	1	.2
		1	1	.3

but at 3^β ,

α	γ	A	C	Pr
1	1	0	0	.4
		1	0	.1
		0	1	.1
		1	1	.4

⁹ k -marginal distribution is the distribution of a subset of k random variables ($k \geq 1$) in a set of $n \geq k$ variables. In Townsend and Schweickert (1989) the property was formulated for 1-marginals of a pair of random variables. The adjective “complete” we use with “marginal selectivity” is to emphasize that we deal with all possible marginals rather than with just 1-marginals.

This means that the diagram of selective influences $(A, B, C) \leftarrow_P (\alpha, \beta, \gamma)$ is ruled out. \square

As pointed out in Section 1, the marginal selectivity property alone is too weak to define selective influences. The example below demonstrates that the property of marginal selectivity does not allow one to treat each of the random variables as preserving its identity in different combinations of “its” factor with other factors.

Example 3.7. Let $\alpha = \{1^\alpha, 2^\alpha\}$, $\beta = \{1^\beta, 2^\beta\}$, and the set of allowable treatments T consist of all four possible combinations of the factor points. Let A and B be Bernoulli variables distributed as shown:

α	β	A	B	Pr
1	1	0	0	.1
		0	1	0
		1	0	0
		1	1	.9

α	β	A	B	Pr
2	1	0	0	0
		0	1	.9
		1	0	.1
		1	1	0

α	β	A	B	Pr
1	2	0	0	.09
		0	1	.01
		1	0	.81
		1	1	.09

α	β	A	B	Pr
2	2	0	0	0
		0	1	.9
		1	0	.1
		1	1	0

Marginal selectivity is satisfied: $\Pr[A(1^\alpha, \cdot) = 0] = 0.1$ and $\Pr[A(2^\alpha, \cdot) = 0] = 0.9$ irrespective of whether the placeholder is replaced with 1^β or 2^β ; and analogously for B . If we assume, however, that this allows us to write $A(1^\alpha)$, $A(2^\alpha)$, $B(1^\beta)$, $B(2^\beta)$ instead of $A(1^\alpha, 1^\beta)$, $A(1^\alpha, 2^\beta)$, etc., we will run into a contradiction. From the tables for $\phi = \{1^\alpha, 1^\beta\}$, $\{2^\alpha, 1^\beta\}$, and $\{2^\alpha, 2^\beta\}$, we can successively conclude $A(1^\alpha) = B(1^\beta)$, $A(2^\alpha) = 1 - B(1^\beta)$, and $A(2^\alpha) = 1 - B(2^\beta)$. But then $A(1^\alpha) = B(2^\beta)$, which contradicts the table for $\phi = \{1^\alpha, 2^\beta\}$, where $A(1^\alpha)$ and $B(2^\beta)$ are stochastically independent and nonsingular. This contradiction proves that the diagram of selective influences $(A, B) \leftarrow_P (\alpha, \beta)$ cannot be inferred from the compliance with marginal selectivity. \square

3.3.3. Invariance under factor-point-specific transformations

Let $(A_1, \dots, A_n) \leftarrow_P (\alpha_1, \dots, \alpha_n)$ and

$$H = \left(H_{x_1^{\alpha_1}}^{\alpha_1}, \dots, H_{x_{k_1}^{\alpha_1}}^{\alpha_1}, \dots, H_{x_1^{\alpha_n}}^{\alpha_n}, \dots, H_{x_{k_n}^{\alpha_n}}^{\alpha_n} \right)$$

be the JDC-vector for $(A_1, \dots, A_n)(\phi)$. Let $F(H)$ be any function that applies to H componentwise and produces a corresponding vector of random variables

$$F(H) = \begin{pmatrix} F(x_1^{\alpha_1}, H_{x_1^{\alpha_1}}^{\alpha_1}), \dots, F(x_{k_1}^{\alpha_1}, H_{x_{k_1}^{\alpha_1}}^{\alpha_1}), \\ \dots, \\ F(x_1^{\alpha_n}, H_{x_1^{\alpha_n}}^{\alpha_n}), \dots, F(x_{k_n}^{\alpha_n}, H_{x_{k_n}^{\alpha_n}}^{\alpha_n}) \end{pmatrix},$$

where we denote by $F(x^\alpha, \cdot)$ the application of F to the component labeled by x^α . Clearly, $F(H)$ possesses a joint distribution and contains one component for each factor point. If we now define a vector of random variables $B(\phi)$ for every treatment $\phi \in T$ as

$$(B_1, \dots, B_n)(\phi) = (F(\phi_{\{\alpha_1\}}, A_1), \dots, F(\phi_{\{\alpha_n\}}, A_n))(\phi),$$

then

$$(B_1, \dots, B_n)(\phi) \sim (F(\phi_{\{\alpha_1\}}, A_1), \dots, F(\phi_{\{\alpha_n\}}, A_n))(\phi),$$

and it follows from JDC that $(B_1, \dots, B_n) \leftarrow_P (\alpha_1, \dots, \alpha_n)$.¹⁰ A function $F(x^{\alpha_i}, \cdot)$ can be referred to as a *factor-point-specific transformation* of the random variable A_i , because the random variable is generally transformed differently for different points of the factor assumed to selectively influence it. We can formulate the property in question by saying that a diagram of selective influences is invariant under all factor-point-specific transformations of the random variables. Note that this includes as a special case transformations which are not factor-point-specific, with

$$F(x_1^{\alpha_i}, \cdot) \equiv \dots \equiv F(x_{k_i}^{\alpha_i}, \cdot) \equiv F(\alpha_i, \cdot).$$

Example 3.8. Let the set-up be the same as in Example 3.7, except for the distributions of (A, B) at the four treatments: we now assume that these distributions are such that $(A, B) \leftarrow_P (\alpha, \beta)$. The tables below show all factor-point-specific transformations $A \rightarrow A^*$ and $B \rightarrow B^*$ at the four treatments, provided that the sets of possible values of A^* and B^* are respectively, $\{\star, \bullet\}$ and $\{\triangleright, \circ\}$, and that at the treatment $\{1^\alpha, 1^\beta\}$ the value 0 of A is mapped into \star and the value 0 of B is mapped into \triangleright .

α	β	$A \rightarrow A^*$	$B \rightarrow B^*$
1	1	0 \rightarrow \star 1 \rightarrow \bullet	0 \rightarrow \triangleright 1 \rightarrow \circ
1	2	0 \rightarrow \star 1 \rightarrow \bullet	0 \rightarrow \circ 1 \rightarrow \triangleright
2	1	0 \rightarrow \bullet 1 \rightarrow \star	0 \rightarrow \triangleright 1 \rightarrow \circ
2	2	0 \rightarrow \bullet 1 \rightarrow \star	0 \rightarrow \circ 1 \rightarrow \triangleright

¹⁰ Since it is possible that $F(x^\alpha, H_{x^\alpha}^{\alpha})$ and $F(y^\alpha, H_{y^\alpha}^{\alpha})$, with $x^\alpha \neq y^\alpha$, have different sets of possible values, strictly speaking, one may need to redefine the functions to ensure that the sets of possible values for $B(\phi)$ is the same for different ϕ . This is, however, not essential (see footnote 6).

α	β	$A \rightarrow A^*$	$B \rightarrow B^*$
1	1	$0 \rightarrow \star$ $1 \rightarrow \bullet$	$0 \rightarrow \triangleright$ $1 \rightarrow \circ$
1	2	$0 \rightarrow \star$ $1 \rightarrow \bullet$	$0 \rightarrow \circ$ $1 \rightarrow \triangleright$
2	1	$0 \rightarrow \star$ $1 \rightarrow \bullet$	$0 \rightarrow \triangleright$ $1 \rightarrow \circ$
2	2	$0 \rightarrow \star$ $1 \rightarrow \bullet$	$0 \rightarrow \circ$ $1 \rightarrow \triangleright$

The possible transformations are restricted to these four because we adhere to our convention that A has the same set of values at all treatments, and the same is true for B . This convention, however, is not essential, and nothing else in the theory prevents one from thinking of A at different treatments as arbitrarily different random variables. With this “relaxed” approach, the following table gives an example of a factor-point-specific transformation:

α	β	$A \rightarrow A^*$	$B \rightarrow B^*$
1	1	$0 \rightarrow 0$ $1 \rightarrow 1$	$0 \rightarrow 0$ $1 \rightarrow 1$
1	2	$0 \rightarrow 0$ $1 \rightarrow 1$	$0 \rightarrow -2$ $1 \rightarrow 3$
2	1	$0 \rightarrow 10$ $1 \rightarrow -20$	$0 \rightarrow 0$ $1 \rightarrow 1$
2	2	$0 \rightarrow 10$ $1 \rightarrow -20$	$0 \rightarrow -2$ $1 \rightarrow 3$

If this is considered undesirable, the variables (A^*, B^*) can be redefined to have $\{-20, 0, 1, 10\}$ and $\{-2, 0, 1, 3\}$ and the respective sets of their possible values, assigning zero probabilities to the values that cannot be attained at a given factor point. \square

This property is of critical importance for construction and use of tests for selective influences, as defined in the next section. A test, generally, lacks the invariance property just formulated: e.g., if the transformation consists in grouping of the original values of random variables, different groupings may result in different outcomes of certain tests, fail or pass. Such a test then can be profitably applied to various factor-point-specific transformations of an original set of random variables, creating thereby in place of a single test a multitude of tests with potentially different outcomes (a single negative outcome ruling out the hypothesis of selective influences).

3.4. General principles for constructing tests for selective influences

3.4.1. Population level tests

Given a set of factors $\{\alpha_1, \dots, \alpha_n\}$, a vector of random variables depending on treatments, $(A_1, \dots, A_n)(\phi)$, and the hypothesis $(A_1, \dots, A_n) \leftarrow_P (\alpha_1, \dots, \alpha_n)$, a *test* for this hypothesis is a

statement \mathfrak{S} relating to each other $(A_1, \dots, A_n)(\phi)$ for different treatments $\phi \in T$ which (a) holds true if $(A_1, \dots, A_n) \leftarrow_P (\alpha_1, \dots, \alpha_n)$, and (b) does not always hold true if this hypothesis is false. A test for a diagram of selective influences therefore is a necessary condition: if the variables $\{(A_1, \dots, A_n)(\phi) : \phi \in T\}$ fail it (i.e., if \mathfrak{S} is false for this set of random variables), we know that the hypothesis $(A_1, \dots, A_n) \leftarrow_P (\alpha_1, \dots, \alpha_n)$ is false. If the statement \mathfrak{S} is always false when $(A_1, \dots, A_n) \not\leftarrow_P (\alpha_1, \dots, \alpha_n)$, the test becomes a *criterion* for selective influences. A test or criterion can be restricted to special classes of random variables (e.g., random variables with finite numbers of values, or multivariate normally distributed at every treatment) and/or factor sets (e.g., 2×2 experimental designs).

The JDC provides a general logic for constructing such tests: we ask whether the hypothetical JDC-vector $H = (H_{x_1^{\alpha_1}}, \dots, H_{x_{k_1}^{\alpha_1}}, \dots, H_{x_1^{\alpha_n}}, \dots, H_{x_{k_n}^{\alpha_n}})$, containing one variable for each factor point of each factor, can be assigned a joint distribution such that its marginals corresponding to the subsets of factor points that form treatments $\phi \in T$ are distributed as $(A_1, \dots, A_n)(\phi)$. Put more succinctly: is there a joint distribution of $(H_{x_1^{\alpha_1}}, \dots, H_{x_{k_1}^{\alpha_1}}, \dots, H_{x_1^{\alpha_n}}, \dots, H_{x_{k_n}^{\alpha_n}})$ with given marginal distributions of the vectors

$$H_\phi = (H_{\phi\{\alpha_1\}}, \dots, H_{\phi\{\alpha_n\}})$$

for all $\phi \in T$?¹¹

Thus, in a study of random variables (A, B) in a 2×2 factorial design, with $\alpha = \{1^\alpha, 2^\alpha\}$, $\beta = \{1^\beta, 2^\beta\}$, and T containing all four logically possible treatments, we consider a hypothetical JDC-vector $(H_{1^\alpha}, H_{2^\alpha}, H_{1^\beta}, H_{2^\beta})$ of which we know the four 2-marginal distributions corresponding to treatments:

$$H_{1^\alpha 1^\beta} = (H_{1^\alpha}, H_{1^\beta}) \sim (A, B) (1^\alpha, 1^\beta),$$

$$H_{1^\alpha 2^\beta} = (H_{1^\alpha}, H_{2^\beta}) \sim (A, B) (1^\alpha, 2^\beta),$$

etc.

Of course, we also know the lower-level marginals, in this case the marginal distributions of H_{1^α} , H_{2^α} , H_{1^β} , and H_{2^β} , but they need not be considered separately as they are determined by the higher-order marginals. The question one poses within the logic of JDC is: can one assign probability densities to different values of $H = (H_{1^\alpha}, H_{2^\alpha}, H_{1^\beta}, H_{2^\beta})$ so that the computed marginal distributions of $(H_{1^\alpha}, H_{1^\beta})$, $(H_{1^\alpha}, H_{2^\beta})$, etc., coincide with the known ones?

If the vector $A = (A_1, \dots, A_n)$ has a finite number of possible values (we may state this without mentioning ϕ because, by

¹¹ Surprisingly, at least for the authors, a slightly less general version of the same problem (the existence of a joint distributions compatible with observable marginals) plays a prominent role in quantum mechanics, in dealing with the quantum entanglement problem (Fine, 1982a-b). We are grateful to Jerome Bussemeyer for bringing this fact to our attention. The parallels with quantum mechanisms will be discussed in a separate publication.

our convention, the set of values does not depend on ϕ), then so does the vector $H = \left(H_{x_1}^{\alpha_1}, \dots, H_{x_{k_1}}^{\alpha_{k_1}}, \dots, H_{x_1}^{\alpha_n}, \dots, H_{x_{k_n}}^{\alpha_{k_n}} \right)$, and the logic of JDC is directly implemented in the *Linear Feasibility Test* introduced in the next section. When the set of values for A is infinite or too large to be handled by the Linear Feasibility Test, one may have to use an indirect approach: computing from the distribution of each H_ϕ certain functionals¹² $g_1(H_\phi), \dots, g_m(H_\phi)$ and constructing a statement

$$\mathfrak{S}(g_1(H_\phi), \dots, g_m(H_\phi) : \phi \in T)$$

relating to each other these functionals for all $\phi \in T$. The statement should be chosen so that it holds true if H possesses a joint distribution, but may be (or, better still, always is) false otherwise.

We illustrate this logic on a simple distance test of the variety introduced in Kujala and Dzhafarov (2008). Assuming that all random variables in (A_1, \dots, A_n) take their values in the set of reals, for each pair of factor points $\{x^\alpha, y^\beta\}$ define

$$Mx^\alpha y^\beta = E \left[|H_{x^\alpha} - H_{y^\beta}| \right],$$

where, for convenience, we write $Mx^\alpha y^\beta$ in place of $M(x^\alpha, y^\beta)$. It can be easily shown that M is a metric on the set H if H possesses a joint distribution for its components. For each treatment ϕ , define the functional

$$g_{\alpha, \beta}(H_\phi) = M\phi_{\{\alpha\}}\phi_{\{\beta\}},$$

whose value can be computed from the known distributions:

$$M\phi_{\{\alpha\}}\phi_{\{\beta\}} = E \left[|A_{\{\alpha\}}(\phi) - A_{\{\beta\}}(\phi)| \right], \quad (11)$$

where $A_{\{\alpha\}}(\phi)$ and $A_{\{\beta\}}(\phi)$ are the random variables in $(A_1, \dots, A_n)(\phi)$ which are supposed to be selectively influenced by α and β , respectively. Due to the marginal selectivity (which we assume to hold because otherwise selective influences have already been ruled out), this quantity is the same for all treatments ϕ which contain the same factor points x^α, y^β of factors α, β . The statement \mathfrak{S} is then as follows: for any (not necessarily pairwise distinct) treatments $\phi^1, \dots, \phi^l \in T$ and any factors $\alpha^1, \dots, \alpha^l \in \Phi$ ($l \geq 3$) such that

$$\alpha_1 \neq \alpha_2 \neq \dots \neq \alpha_{l-1} \neq \alpha_l \neq \alpha_1, \quad (12)$$

and

$$\phi_{\{\alpha^1\}}^1 = \phi_{\{\alpha^1\}}^2, \dots, \phi_{\{\alpha^{l-1}\}}^{l-1} = \phi_{\{\alpha^{l-1}\}}^l, \phi_{\{\alpha^l\}}^l = \phi_{\{\alpha^1\}}^1, \quad (13)$$

we should have

$$g_{\alpha^1, \alpha^l}(H_{\phi^1}) \leq g_{\alpha^1, \alpha^2}(H_{\phi^2}) + \dots + g_{\alpha^{l-1}, \alpha^l}(H_{\phi^l}). \quad (14)$$

¹² A functional $g(X)$ is a function mapping each random variable X from some set of random variables into, typically, a real or complex number (more generally, an element of a certain “standard” set). A typical example of a functional is the expected value $E[X]$.

The truth of \mathfrak{S} for H with jointly distributed components follows from the triangle inequality for M . The inequality may very well be violated when the components of H do not possess a joint distribution (i.e., when the hypothesis of selective influences is false).

Example 3.9. To apply this test to Example 3.7, we make use of the property that if $(A, B) \leftarrow_P (\alpha, \beta)$ then $(A^*, B^*) \leftarrow_P (\alpha, \beta)$ for any factor-point-specific transformations (A^*, B^*) of (A, B) . Let us put $B^* = B$ and

$$A^* = \begin{cases} A & \text{if } \phi_{\{\alpha\}} = 1^\alpha, \\ 1 - A & \text{if } \phi_{\{\alpha\}} = 2^\alpha. \end{cases}$$

This yields the distributions

α	β	A^*	B^*	Pr
1	1	0	0	.1
		0	1	0
		1	0	0
		1	1	.9
α	β	A^*	B^*	Pr
2	1	1	0	0
		1	1	.9
		0	0	.1
		0	1	0

α	β	A^*	B^*	Pr
1	2	0	0	.09
		0	1	.01
		1	0	.81
		1	1	.09
α	β	A^*	B^*	Pr
2	2	1	0	0
		1	1	.9
		0	0	.1
		0	1	0

It is easy to check that

$$M1^\alpha 1^\beta = E \left[|A(1^\alpha, 1^\beta) - B(1^\alpha, 1^\beta)| \right] = 0,$$

$$M1^\alpha 2^\beta = E \left[|A(1^\alpha, 2^\beta) - B(1^\alpha, 2^\beta)| \right] = 0.82,$$

$$M2^\alpha 1^\beta = E \left[|A(2^\alpha, 1^\beta) - B(2^\alpha, 1^\beta)| \right] = 0,$$

$$M2^\alpha 2^\beta = E \left[|A(2^\alpha, 2^\beta) - B(2^\alpha, 2^\beta)| \right] = 0.$$

Since

$$0.82 = M1^\alpha 2^\beta > M1^\alpha 1^\beta + M2^\alpha 1^\beta + M2^\alpha 2^\beta = 0,$$

the triangle inequality is violated, rejecting thereby the hypothesis $(A^*, B^*) \leftarrow_P (\alpha, \beta)$, hence also the hypothesis $(A, B) \leftarrow_P (\alpha, \beta)$. \square

3.4.2. Sample-level tests

Although this paper is not concerned with statistical questions, it may be useful to outline the general logic of constructing a sample-level test corresponding to a population-level one. Analytic procedures and asymptotic approximations have to be different for different tests, but if the population-level test can be computed efficiently, the following Monte-Carlo procedure is always applicable.

1. For each of the random variables A_1, \dots, A_n , if it has more than a finite number of values (or has too many values, even if finite), we discretize it in the conventional way, by forming successive adjacent intervals and replacing each of them with its midpoint. Continue to denote the discretized random variables A_1, \dots, A_n .
2. We now have sample proportions $\hat{\Pr}[(A_1 = a_1, \dots, A_n = a_n) (x_1^{\alpha_1}, \dots, x_n^{\alpha_n})]$, where a_1, \dots, a_n are possible values of the corresponding random variables A_1, \dots, A_n .
3. For each treatment, we form a confidence region of possible probabilities $\Pr[(A_1 = a_1, \dots, A_n = a_n) (x_1^{\alpha_1}, \dots, x_n^{\alpha_n})]$ for a given set of estimates, at a given level of a familywise confidence level for the Cartesian product of these confidence regions, with an appropriately adopted convention on how this familywise confidence is computed (glossing over a controversial issue).
4. The hypothesis of selective influences is retained or rejected according as the combined confidence region contains or does not contain a point (a set of joint probabilities) which passes the population test in question. (Gradualized versions of this procedure are possible, when each point in the space of population-level probabilities is taken with the weight proportional to its likelihood.)

Instead of a confidence region of multivariate distributions based on a discretization, one can also generate confidence regions of distributions belonging to a specified class, say, multivariate normal ones.

Resampling techniques is another obvious approach, although the results will generally depend on one's often arbitrary choice of the resampling procedure. One simple choice is the permutation test in which the joint sample proportions $\hat{\Pr}[A_1 = a_1, \dots, A_n = a_n]$ obtained at different treatments (and treated as probabilities) are randomly assigned to the treatments ϕ . If the initial, observed assignment passes a test, while the proportion of the permuted assignments which pass the test is sufficiently small, the hypothesis of selective influences is considered supported.

4. LINEAR FEASIBILITY TEST

In this section we assume that each random variable $A_i(\phi)$ in $(A_1, \dots, A_n)(\phi)$ has a finite number m_i of possible values a_{i1}, \dots, a_{im_i} . It is arguably the most important special case both because it is ubiquitous in psychological theories and because in all other cases random variables can be discretized into finite number of categories. We are interested in establishing the truth or falsity of the diagram of selective influences (8), where each factor α_i in $(\alpha_1, \dots, \alpha_n)$ contains k_i factor points. The *Linear Feasibility Test* to be described is a direct application of JDC to this situation,¹³ furnishing both a necessary

and sufficient condition for the diagram of selective influences $(A_1, \dots, A_n) \leftarrow_P (\alpha_1, \dots, \alpha_n)$.

In the hypothetical JDC-vector

$$H = \left(H_{x_1^{\alpha_1}}, \dots, H_{x_{k_1}^{\alpha_1}}, \dots, H_{x_1^{\alpha_n}}, \dots, H_{x_{k_n}^{\alpha_n}} \right),$$

since we assume that

$$H_{x_j^{\alpha_i}} \sim A_i(\phi)$$

for any $x_j^{\alpha_i}$ and any treatment ϕ containing $x_j^{\alpha_i}$, we know that the set of possible values for the random variable $H_{x_j^{\alpha_i}}$ is $\{a_{i1}, \dots, a_{im_i}\}$, irrespective of x_j . Denote

$$\begin{aligned} & \Pr \left[(A_1 = a_{1l_1}, \dots, A_n = a_{nl_n}) (x_{\lambda_1}^{\alpha_1}, \dots, x_{\lambda_n}^{\alpha_n}) \right] \\ &= P \left(\overbrace{l_1, \dots, l_n}^{\text{for r.v.s}}; \overbrace{\lambda_1, \dots, \lambda_n}^{\text{for factor points}} \right), \end{aligned} \quad (15)$$

where $l_i \in \{1, \dots, m_i\}$ and $\lambda_i \in \{1, \dots, k_i\}$ for $i = 1, \dots, n$ ("r.v.s" abbreviates "random variables"). Denote

$$\begin{aligned} & \Pr \left[\begin{array}{c} H_{x_1^{\alpha_1}} = a_{1l_{11}}, \dots, H_{x_{k_1}^{\alpha_1}} = a_{1l_{1k_1}}, \\ \dots, \\ H_{x_1^{\alpha_n}} = a_{nl_{n1}}, \dots, H_{x_{k_n}^{\alpha_n}} = a_{nl_{nk_n}} \end{array} \right] \\ &= Q \left(\overbrace{l_{11}, \dots, l_{1k_1}}^{\text{for } A_1}, \dots, \overbrace{l_{n1}, \dots, l_{nk_n}}^{\text{for } A_n} \right), \end{aligned} \quad (16)$$

where $l_{ij} \in \{1, \dots, m_i\}$ for $i = 1, \dots, n$. This gives us $m_1^{k_1} \times \dots \times m_n^{k_n}$ Q -probabilities. A required joint distribution for the JDC-vector H exists if and only if these probabilities can be found subject to $m_1^{k_1} \times \dots \times m_n^{k_n}$ nonnegativity constraints

$$Q(l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) \geq 0, \quad (17)$$

and (denoting by n_T the number of treatments in T) $n_T \times m_1 \times \dots \times m_n$ linear equations

$$\begin{aligned} & \sum Q(l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) \\ &= P(l_1, \dots, l_n; \lambda_1, \dots, \lambda_n), \end{aligned} \quad (18)$$

where the summation is across all possible values of the set

$$\{l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}\} - \{l_{1\lambda_1}, \dots, l_{n\lambda_n}\},$$

while

$$l_{1\lambda_1} = l_1, \dots, l_{n\lambda_n} = l_n.$$

¹³ In reference to footnote 11, this test has been proposed in the context of dealing with multiple-particle multiple-measurement quantum entanglement situations

by Werner & Wolf (2001a, b) and Basoalto & Percival (2003).

Selective influences hold if and only if the system of these linear equalities with the nonnegativity constraints is *feasible* (i.e., has a solution). This is a typical *linear programming* problem (see, e.g., Webster, 1994, Ch. 4).¹⁴ Many standard statistical and mathematical packages can handle this problem.

Note that the maximal value for n_T is $n_T = k_1 \times \dots \times k_n$, whence the maximal number of linear equations is $(m_1 k_1) \times \dots \times (m_n k_n)$. Since $m_i k_i \leq m_i^{k_i}$ (assuming $m_i, k_i \geq 2$), with the equality only achieved at $k_i = m_i = 2$, the system of linear equations is always underdetermined. In fact, the system of equations is underdetermined even if $k_i = m_i = 2$ for all $i = 1, \dots, n$, because of the obvious linear dependences among the equations.

Example 4.1. Let $\alpha = \{1^\alpha, 2^\alpha\}$, $\beta = \{1^\beta, 2^\beta\}$, and the set of allowable treatments T consist of all four possible combinations of the factor points. Let A and B be Bernoulli variables distributed as shown:

α	β	A	B	Pr
1	1	0	0	.140
		0	1	.360
		1	0	.360
		1	1	.140
α	β	A	B	Pr
2	1	0	0	.189
		0	1	.311
		1	0	.311
		1	1	.189

Marginal selectivity here is satisfied trivially: all marginal probabilities are equal 0.5, for all treatments. The linear programming routine of MathematicaTM (using the interior point algorithm) shows that the linear equations (18) have nonnegative solutions corresponding to the JDC-vector

H_{1^α}	H_{2^α}	H_{1^β}	H_{2^β}	Pr
0	0	0	0	.02708610
0	0	0	1	.00239295
0	0	1	0	.16689300
0	0	1	1	.03358610
0	1	0	0	.00197965
0	1	0	1	.10854100
0	1	1	0	.00204128
0	1	1	1	.15748000

This proves that in this case we do have $(A, B) \leftarrow (\alpha, \beta)$. \square

Example 4.2. In the previous example, let us change the distributions of (A, B) to the following:

α	β	A	B	Pr
1	1	0	0	.450
		0	1	.050
		1	0	.050
		1	1	.450
α	β	A	B	Pr
2	1	0	0	.170
		0	1	.330
		1	0	.330
		1	1	.170

Once again, marginal selectivity is satisfied trivially, as all marginal probabilities are 0.5, for all treatments. The linear programming routine of MathematicaTM, however, shows that the linear equations (18) have no nonnegative solutions. This excludes the existence of a JDC-vector for this situations, ruling out thereby the possibility of $(A, B) \leftarrow (\alpha, \beta)$. \square

Since the Linear Feasibility Test is both a necessary and sufficient condition for selective influences, if it is passed for $(A_1, \dots, A_n)(\phi)$, it is guaranteed to be passed following any factor-point-specific transformations of these random outputs. All such transformations in the case of discrete random variables can be described as combinations of renamings (factor-point specific ones) and augmentations (grouping of some values together). In fact, a result of the Linear Feasibility Test simply does not depend on the values of the random variables involved, only their probabilities matter. Therefore a renaming, such as in Example 3.8, will not change anything in the system of linear equations and inequalities (17)-(18). An example of augmentation (or “coarsening”) will be redefining A and B , each having possible values 1, 2, 3, 4, into binary variables

$$A^*(\phi) = \begin{cases} 0 & \text{if } A(\phi) = 1, 2, \\ 1 & \text{if } A(\phi) = 3, 4, \end{cases} \quad B^*(\phi) = \begin{cases} 0 & \text{if } B(\phi) = 1, 2, 3, \\ 1 & \text{if } B(\phi) = 4. \end{cases}$$

It is clear that any such an augmentation amounts to replacing some of the equations in (18) with their sums. Therefore, if the original system has a solution, so will also the system after such replacements.

The same reasoning applies to one’s redefining the factors by grouping together some of the factor points: e.g., redefining $\alpha = \{1^\alpha, 2^\alpha, 3^\alpha\}$ into

$$\alpha^* = \left\{ \{1^\alpha, 2^\alpha\}^{\alpha^*}, \{3^\alpha\}^{\alpha^*} \right\} = \left\{ 1^{\alpha^*}, 2^{\alpha^*} \right\}.$$

This change will amount to replacing by their sum any two equations whose right hand sides correspond to identical vectors $(l_1, \dots, l_n; \lambda_1, \dots, \lambda_n)$ except for the factor point for α being 1 in one of them and 2 in another.

Summarizing, the Linear Feasibility Test cannot reject selective influences on a coarser level of representation (for random variables and/or factors) and uphold it on a finer level (although the reverse, obviously, can happen).

If the random variables involved have more than finite number of values and/or the factors consist of more than finite number of

¹⁴ More precisely, this is a linear programming task in the standard form and with a dummy objective function (e.g., a linear combination with zero coefficients).

factor points, or if these numbers, though finite, are too large to handle the ensuing linear programming problem, then the Linear Feasibility Test can still be used after the values of the random variables and/or factors have been appropriately grouped. The Linear Feasibility Test then becomes only a necessary condition for selective influences, and its results will generally be different for different (non-nested) groupings.

Example 4.3. Consider the hypothesis $(A, B) \leftrightarrow (\alpha, \beta)$ with the factors having a finite number of factor points each, and A and B being response times. To use the Linear Feasibility Test, one can transform the random variable A as, say,

$$A^*(\phi) = \begin{cases} 1 & \text{if } A(\phi) \leq a_{1/4}(\phi), \\ 2 & \text{if } a_{1/4}(\phi) < A(\phi) \leq a_{1/2}(\phi), \\ 3 & \text{if } a_{1/2}(\phi) < A(\phi) \leq a_{3/4}(\phi), \\ 4 & \text{if } A(\phi) > a_{3/4}(\phi), \end{cases}$$

and transform B as

$$B^*(\phi) = \begin{cases} 1 & \text{if } B(\phi) \leq b_{1/2}(\phi), \\ 2 & \text{if } B(\phi) > b_{1/2}(\phi), \end{cases}$$

where $a_p(\phi)$ and $b_p(\phi)$ designate the p th quantiles of, respectively $A(\phi)$ and $B(\phi)$. The initial hypothesis now is reformulated as $(A^*, B^*) \leftrightarrow (\alpha, \beta)$, with the understanding that if it is rejected then the initial hypothesis will be rejected too (a necessary condition only). The Linear Feasibility test will now be applied to distributions of the form

α	β	A	B	Pr
x	y	1	1	p_{11}
		1	2	p_{12}
	\vdots	\vdots	\vdots	\vdots
		4	1	p_{41}
		4	2	p_{42}

where the marginals for A are constrained to 0.25 and the marginals for B to 0.5, for all treatments $\{x^\alpha, y^\beta\}$, yielding a trivial compliance with marginal selectivity. Note that the test may very well uphold $(A^*, B^*) \leftrightarrow (\alpha, \beta)$ even if marginal selectivity is violated for $(A, B)(\phi)$ (e.g., if the quantiles $a_p(x^\alpha, y^\beta)$ change as a function of y^β). \square

Sample level problems do not seem to present a serious difficulty. The general approach mentioned in Section 3.4.2 is facilitated by the following consideration. If a system of linear equations and inequalities has an “interior” solution (one for which all inequalities are satisfied in the strict form, which in our case means that the solution contains no zeros), then the solution is stable with respect to sufficiently small perturbations of its coefficients. In our case, this means that if an interior solution exists for population-level values of $P(l_1, \dots, l_n; \lambda_1, \dots, \lambda_n)$, and if the sample estimates of the latter are sufficiently close to the population values, then the system will also have a solution for sample estimates. By the same token, if no solution exists for the population-level values of $P(l_1, \dots, l_n; \lambda_1, \dots, \lambda_n)$, then no

solution will be found for sufficiently close to them sample estimates. The only unstable situation exists if solutions exist on the hypothetical population level (i.e., the selectiveness of influences is satisfied), but they are all non-interior (contain zeros).

Remark 4.4. The question arises: how restrictive is the condition of selective influences within the class of distributions satisfying marginal selectivity? We do not know anything close to a complete answer to this question, but simulations show that selectivity of influence is not overly restrictive with respect to marginal selectivity. Thus, if $k_i = m_i = 2$ for $i = 1, 2$, and if we constrain all marginal probabilities to 0.5 and pick $P(1, 1; 1, 1), P(1, 1; 1, 2), P(1, 1; 2, 1), P(1, 1; 2, 2)$ from four independent uniform distributions between 0 and 0.5, the probability of “randomly” obtaining selective influences is about 0.67. If $k_i = m_i = 2$ for $i = 1, 2, 3$, and we constrain all 2-marginal probabilities to 0.25, the analogous probability is about 0.10.

5. DISTANCE-TYPE TESTS

5.1. General theory

First, we establish the general terminology related to distance-type functions. Given a set \mathcal{R} , a function $d : \mathcal{R} \times \mathcal{R} \rightarrow [0, \infty]$ is a *premetric* if $d(x, x) = 0$. The inclusion of the possibility $d(x, y) = \infty$ usually adds the qualifier “extended” (in this case, extended premetric), but we will omit it for brevity. A premetric that satisfies the triangle inequality,

$$d(x, z) \leq d(x, y) + d(y, z),$$

for any $x, y, z \in \mathcal{R}$, is a *pseudo-quasi-metric* (*p.q.-metric*, for short). A p.q.-metric which is symmetric,

$$d(x, y) = d(y, x),$$

for all $x, y \in \mathcal{R}$, is a *pseudometric*. A p.q.-metric such that

$$x \neq y \implies d(x, y) > 0$$

(equivalently, $d(x, y) = 0$ if and only if $x = y$) is a *quasimetric*. A p.q.-metric which is simultaneously a quasimetric and a pseudometric is a *conventional* (symmetric) *metric*. The words “metric” and “distance” can be used interchangeably: so one can speak of conventional (symmetric) distances, pseudodistances, quasidistances, and p.q.-distances.¹⁵

We are interested in the situation when \mathcal{R} is a set of jointly distributed random variables (discrete, continuous, or mixed), with the intent to apply a distance-type function definable on such an \mathcal{R} to the JDC-vector H of random variables for the diagram of selective influences (8). The random variables $A(\phi) = (A_1, \dots, A_n)(\phi)$, the factors $\Phi = \{\alpha_1, \dots, \alpha_n\}$, and the set of

¹⁵ The terminology adopted in this paper is conventional but not universal. In particular, the term “metric” or “distance” is sometimes used to mean pseudometric. In the context of Finsler geometry and the dissimilarity cumulation theory (Dzhafarov, 2010) the term “metric” is used to designate quasimetric with an additional property of being “symmetric in the small.”

treatments T are defined as above. The main property we are concerned with is the triangle inequality, that is, it is typically sufficient to know that the distance-type function we are dealing with is a p.q.-metric.

The function (11) considered in Section 3.4.1 serves as an introductory example of a metric on which one can base a test for selective influences. As a simple example of using a p.q.-metric which is not a conventional metric (in fact, not even a pseudo-metric or quasimetric), consider the following. Let the elements of \mathcal{R} be binary random variables, with values $\{1, 2\}$. Define, for any $A_1, \dots, A_p, B_1, \dots, B_q \in \mathcal{R}$,

$$P^{(2)}[(A_1, \dots, A_p)(B_1, \dots, B_q)] = \Pr \left[\begin{array}{l} A_i = 1 \text{ for } i = 1, \dots, p, \\ B_j = 2 \text{ for } j = 1, \dots, q \end{array} \right].$$

The parentheses may be dropped around singletons, in particular,

$$\Pr[A = 1, B = 2] = P^{(2)}[(A)(B)] = P^{(2)}[AB].$$

The latter is clearly a premetric: $P^{(2)}$ is nonnegative, and $P^{(2)}[RR] = 0$, for any $R \in \mathcal{R}$. To prove the triangle inequality,

$$P^{(2)}[R_1 R_2] \leq P^{(2)}[RR_2] + P^{(2)}[R_1 R],$$

for any $R_1, R_2, R \in \mathcal{R}$, observe that

$$P^{(2)}[R_1 R_2] = P^{(2)}[(R_1, R) R_2] + P^{(2)}[R_1 (R_2, R)],$$

$$P^{(2)}[RR_2] = P^{(2)}[(R_1, R) R_2] + P^{(2)}[R (R_1, R_2)],$$

$$P^{(2)}[R_1 R] = P^{(2)}[(R_1, R_2) R] + P^{(2)}[R_1 (R_2, R)],$$

whence

$$\begin{aligned} & P^{(2)}[RR_2] + P^{(2)}[R_1 R] - P^{(2)}[R_1 R_2] \\ &= P^{(2)}[R (R_1, R_2)] + P^{(2)}[(R_1, R_2) R] \geq 0. \end{aligned}$$

Note that $P^{(2)}$ is not a pseudometric because generally

$$\begin{aligned} P^{(2)}[R_1 R_2] &= \Pr[R_1 = 1, R_2 = 2] \\ &\neq \Pr[R_2 = 1, R_1 = 2] = P^{(2)}[R_2 R_1]. \end{aligned}$$

Nor is $P^{(2)}$ a quasimetric because it may very well happen that $R_1 \neq R_2$ but

$$P^{(2)}[R_1 R_2] = \Pr[R_1 = 1, R_2 = 2] = 0.$$

To use this p.q.-metric for our purposes: each random variable H_{x^α} in the hypothetical JDC-vector H has a set of possible values \mathcal{A}_α , in which we choose and fix a measurable subset $\mathcal{A}_{x^\alpha}^+$ and its complement $\mathcal{A}_{x^\alpha}^-$. Note that \mathcal{A}_α is the same for all factor points of the factor α (and coincides with the spectrum of the random variable in the diagram (6) which is supposed to be selectively influenced by α). Transform each H_{x^α} as

$$R_{x^\alpha} = \begin{cases} 1 & \text{if } H_{x^\alpha} \in \mathcal{A}_{x^\alpha}^-, \\ 2 & \text{if } H_{x^\alpha} \in \mathcal{A}_{x^\alpha}^+, \end{cases} \quad (19)$$

and define, for each pair of factor points x^α, y^β ,

$$Dx^\alpha y^\beta = P^{(2)}[R_{x^\alpha} R_{y^\beta}]. \quad (20)$$

Here, once again (see Section 3.4.1), we write $x^\alpha y^\beta$ in place of (x^α, y^β) . This time we are going to formalize this notation as part the following general convention: any *chain* (a finite sequence) of factor points will be written as a *string of symbols*, without commas and parentheses, such as $x_1^{\alpha_1} \dots x_l^{\alpha_l}, x^\alpha y^\beta z^\gamma$, etc.

The value of $Dx^\alpha y^\beta$ is computable for any $x^\alpha y^\beta$ which is part of a treatment $\phi \in T$. The test therefore consists in checking whether

$$Dx_1^{\alpha_1} x_l^{\alpha_l} \leq Dx_1^{\alpha_1} x_2^{\alpha_2} + Dx_2^{\alpha_2} x_3^{\alpha_3} + \dots + Dx_{l-1}^{\alpha_{l-1}} x_l^{\alpha_l} \quad (21)$$

for any chain of factor points $x_1^{\alpha_1} \dots x_l^{\alpha_l}$ ($l \geq 3$) satisfying (12) and such that for some treatments $\phi^{(1)}, \dots, \phi^{(l)} \in T$ (not necessarily pairwise distinct),

$$\{x_1^{\alpha_1}, x_l^{\alpha_l}\} \subset \phi^{(1)}, \{x_1^{\alpha_1}, x_2^{\alpha_2}\} \subset \phi^{(2)}, \dots, \{x_{l-1}^{\alpha_{l-1}}, x_l^{\alpha_l}\} \subset \phi^{(l)}. \quad (22)$$

Note that this is just another way of writing (13)-(14). If the test is failed (i.e., the inequality is violated) for at least one such sequence of factor points, then the hypothesis $(A_1, \dots, A_n) \leftarrow \mathcal{P}(\alpha_1, \dots, \alpha_n)$ is rejected. In the following we will refer to any sequence of factor points $x_1^{\alpha_1} \dots x_l^{\alpha_l}$ ($l \geq 3$) subject to (12) and (22) as a *treatment-realizable chain*.

Example 5.1. Let $\alpha = \{1^\alpha, 2^\alpha\}$, $\beta = \{1^\beta, 2^\beta\}$, and the set of allowable treatments T consist of all four possible combinations of the factor points. Let (A, B) be bivariate normally distributed at every treatment ϕ , with standard normal marginals and with correlations

$$\rho(x^\alpha, y^\beta) = \begin{cases} -.9 & \text{at } \{x^\alpha, y^\beta\} = \{1^\alpha, 1^\beta\}, \\ +.9 & \text{at } \{x^\alpha, y^\beta\} = \{1^\alpha, 2^\beta\}, \\ +.9 & \text{at } \{x^\alpha, y^\beta\} = \{2^\alpha, 1^\beta\}, \\ -.1 & \text{at } \{x^\alpha, y^\beta\} = \{2^\alpha, 2^\beta\}. \end{cases}$$

We form variables

$$A^*(\phi) = \begin{cases} 1 & \text{if } A(\phi) \leq 0, \\ 2 & \text{if } A(\phi) > 0, \end{cases} \quad B^*(\phi) = \begin{cases} 1 & \text{if } B(\phi) \leq 0, \\ 2 & \text{if } B(\phi) > 0, \end{cases}$$

with all marginals obviously constrained to 0.5, for all treatments. The joint distributions are computed to be

α	β	A^*	B^*	Pr
1	1	1	1	...
		1	2	.428217
		2	1	...
		2	2	...
α	β	A^*	B^*	Pr
2	1	1	1	...
		1	2	.0717831
		2	1	...
		2	2	...

α	β	A^*	B^*	Pr
1	2	1	1	...
		1	2	.0717831
		2	1	...
		2	2	...
α	β	A^*	B^*	Pr
2	2	1	1	...
		1	2	.265942
		2	1	...
		2	2	...

where for each treatment ϕ we only show the probabilities $\Pr[A^* = 1, B^* = 2] = P^{(2)}[A^*B^*]$, other probabilities being irrelevant for our computations. Since $\{1^\alpha, 1^\beta\}$, $\{1^\alpha, 2^\beta\}$, $\{2^\alpha, 2^\beta\}$, and $\{2^\alpha, 1^\beta\}$ are all allowable treatment, $1^\alpha 2^\beta 2^\alpha 1^\beta$ is a treatment-realizable chain. We can put therefore

$$Dx^\alpha y^\beta = P^{(2)}[A^*(x^\alpha, y^\beta)B^*(x^\alpha, y^\beta)]$$

and observe that

$$.428217 = D1^\alpha 1^\beta > D1^\alpha 2^\beta + D2^\alpha 2^\beta + D2^\alpha 1^\beta = 0.409508.$$

This violation of the chain inequality rules out $(A, B) \leftrightarrow (\alpha, \beta)$. \square

The formulation of the test (21), subject to (12) and (22), is valid for any p.q.-metric D imposed on the hypothetical JDC-vector H for the diagram (8). It turns out, however, that using all possible treatment-realizable chains $x_1^{\alpha_1} \dots x_l^{\alpha_l}$ of factor points would be redundant, in view of the lemma below. For its formulation we need an additional concept. A treatment-realizable chain $x_1^{\alpha_1} \dots x_l^{\alpha_l}$ ($l \geq 3$) is called *irreducible* if

1. the only nonempty subsets thereof that are subsets of treatments are the pairs listed in (22), and
2. no factor point in it occurs more than once.

Thus, a triadic treatment-realizable chain $x^\alpha y^\beta z^\gamma$ is irreducible if and only if there is no treatment ϕ that includes $\{x^\alpha, y^\beta, z^\gamma\}$. Tetradic treatment-realizable chains of the form $x^\alpha y^\beta u^\alpha v^\beta$ are irreducible if and only if $x^\alpha \neq u^\alpha$ and $y^\beta \neq v^\beta$.

Theorem 5.2 (Distance-type Tests). *Given a p.q.-metric D on the hypothetical JDC-vector H for the diagram (8), the inequality (21) is satisfied for all treatment-realizable chains if and only if this inequality holds for all irreducible chains.*

This theorem is an immediate consequence of Lemma A.11 in the appendix, where it is proved for a general set-up involving arbitrary sets of random entities and factors.

Note that if T includes all possible combinations of factor points, $T = \alpha_1 \times \dots \times \alpha_m$ (“completely crossed design”), then the condition of treatment-realizability is equivalent to (12). In this situation any set of factor points belonging to pairwise different factors (e.g., $\{x^\alpha, y^\beta\}$, or $\{x^\alpha, y^\beta, z^\gamma\}$ with $\alpha \neq \beta \neq \gamma \neq \alpha$) belongs to some treatment, whence an irreducible chain cannot contain factor points of more than two distinct factors: they must all be of the form $x_1^\alpha x_2^\beta x_3^\alpha x_4^\beta \dots x_{2k-1}^\alpha x_{2k}^\beta$ ($\alpha \neq \beta$). It is easy to see, however, that if $k > 2$, each of the subsets $\{x_1^\alpha, x_4^\beta\}$ and $\{x_2^\beta, x_5^\alpha\}$ belongs to a treatment. It follows that all irreducible chains in a completely crossed design are of the form $x^\alpha y^\beta u^\alpha v^\beta$, with $\alpha \neq \beta$, $x^\alpha \neq u^\alpha$ and $y^\beta \neq v^\beta$.

Theorem 5.3 (Distance-type Tests for Completely Crossed Designs). *If the set of treatments T consists of all possible combinations of factor points, then the inequality (21) is satisfied for all treatment-realizable sequences of factor points if and only if this inequality holds for all tetradic sequences of the form $x^\alpha y^\beta u^\alpha v^\beta$, with $\alpha \neq \beta$, $x^\alpha \neq u^\alpha$ and $y^\beta \neq v^\beta$.*

This formulation is given in Dzhaferov and Kujala (2010), although there it is unnecessarily confined to metrics of a special kind, denoted $M^{(p)}$ below.

5.2. Classes of p.q.-metrics

Let us consider some classes of p.q.-metrics that can be used for distance-type tests. We do not attempt a systematization or maximal generality, our goals being to show the reader how broad the spectrum of the usable p.q.-metrics is, and how easy it is to generate new ones.

5.2.1. Minkowski-type metrics

These are (conventional, symmetric) metrics of the type

$$M^{(p)}(A, B) = \begin{cases} \sqrt[p]{E[|A - B|^p]} & \text{for } 1 \leq p < \infty, \\ \text{ess sup } |A - B| & \text{for } p = \infty, \end{cases} \quad (23)$$

where

$$\text{ess sup } |A - B| = \inf \{v : \Pr[|A - B| \leq v] = 1\}.$$

In the context of selective influences these metrics have been introduced in Kujala and Dzhaferov (2008) and further analyzed in Dzhaferov and Kujala (2010). The metric M discussed in Section 3.4.1 is a special case ($p = 1$). An important property of $M^{(p)}$ is that the result of an $M^{(p)}$ -based distance-type test is not invariant with respect to factor-point-specific transformations of the random variables. This allows one to conduct an infinity of different tests on one and the same $A(\phi) = (A_1, \dots, A_n)(\phi)$. For numerous examples of how the test works see Kujala and Dzhaferov (2008) and Dzhaferov and Kujala (2010).

5.2.2. Classification p.q.-metrics

Classification p.q.-metrics are the p.q.-metrics defined through the p.q.-metric $P^{(2)}$ by (20), following a transformation (19). The general definition is that for each random variable X in a set of jointly distributed random variables \mathcal{R} we designate two complementary events E_X^- and E_X^+ , and put

$$D_C(A, B) = \Pr[E_A^- \& E_B^+].$$

The results of a D_C -based distance-type test for selective influences depend on the choice of the events E_X^\pm , so different choices would lead to different tests for one and the same $A(\phi) = (A_1, \dots, A_n)(\phi)$. See Example 5.1 for an illustration.

To the best of our knowledge this interesting p.q.-metric was not previously considered in mathematics. One standard way to generalize it (see the principles of constructing derivative metrics in Section 5.2.4 below) is to make the set of events $\{E_X^\pm : X \in \mathcal{R}\}$ a random entity. In the special case when all random variables in \mathcal{R} take their values in the set of real numbers, and E_X^\pm for each $X \in \mathcal{R}$ is defined by $X \geq v$, the “randomization”

of $\{E_X^+ : X \in \mathcal{R}\}$ reduces to that of v . The p.q.-metric then becomes

$$D_S(A, B) = \Pr[A \leq V < B]$$

where V is a random variable. An additively symmetrized (i.e., pseudometric) version of this p.q.-metric, $D_S(A, B) + D_S(B, A)$, was introduced in Taylor (1984, 1985) under the name “separation (pseudo)metric,” and shown to be a conventional metric if V is chosen stochastically independent of all random variables in \mathcal{R} .

5.2.3. Information-based p.q.-metric

Let the jointly distributed random variables constituting the set \mathcal{R} be all discrete. Perhaps the simplest information-based p.q.-metric is

$$h(A|B) = - \sum_{a,b} p_{AB}(a, b) \log \frac{p_{AB}(a, b)}{p_B(b)},$$

with the conventions $0 \log \frac{0}{0} = 0 \log 0 = 0$. This function is called *conditional entropy*. The identity $h(A|A) = 0$ is obvious, and the triangle inequality,

$$h(A|B) \leq h(A|C) + h(C|B),$$

follows from the standard information theory (in)equalities,

$$h(A|B) \leq h(A, C|B),$$

$$h(A, C|B) = h(A|C, B) + h(C|B),$$

and

$$h(A|C, B) \leq h(A|C).$$

Note that the test of selectiveness based on $h(A, B)$ (and any other information-based measure) is invariant with respect to all bijective transformations of the variables.

The additively symmetrized (i.e., pseudometric) version of this p.q.-metric, $h(A|B) + h(B|A)$ is well-known (Cover & Thomas, 1990). Normalized versions of $h(A|B)$ are also of interest, for instance,

$$h_N(A|B) = \frac{2h(A|B)}{h(A, B)},$$

where

$$h(A, B) = - \sum_{a,b} p_{AB}(a, b) \log p_{AB}(a, b),$$

the *joint entropy* of A and B ; $h_N(A|B)$ is bound between 0 (attained when A is a bijective transformation of B) and 1 (when A and B are independent). A proof of the triangle inequality for h_N can be found in Kraskov et al. (2003), as part of their proof that $\frac{1}{2} [h_N(A|B) + h_N(B|A)]$ is a pseudometric.

5.2.4. Constructing p.q.-metrics from other p.q.-metrics

There are numerous ways of creating new p.q.-metrics from the ones mentioned above, or from ones taken from outside probabilistic context. Thus, if d is a p.q.-metric on a set S , then, for any space \mathcal{R} of jointly distributed random variables taking their values in S ,

$$D(A, B) = E[d(A, B)], \quad A, B \in \mathcal{R},$$

is a p.q.-metric on \mathcal{R} . This follows from the fact that expectation E preserves inequalities and equalities identically satisfied for all possible realizations of the arguments. Thus, the distance $M(A, B) = E[|A - B|]$ of Section 3.4.1 trivially obtains from the metric $d(a, b) = |a - b|$ on reals. In the same way one obtains the well-known Fréchet distance

$$F(A, B) = E \left[\frac{|A - B|}{1 + |A - B|} \right].$$

Below we present an incomplete list of transformations which, given a p.q.-metric (quasimetric, pseudometric, conventional metric) D on a space \mathcal{R} of jointly distributed random variables produces a new p.q.-metric (respectively, quasimetric, pseudometric, or conventional metric) on the same space. The proofs are trivial or well-known, so we omit them. The arrows \implies should be read “can be transformed into.”

1. $D \implies D^q$ ($q < 1$). In this way, for example, we can obtain metrics

$$M^{(p,q)}(A, B) = \begin{cases} (E[|A - B|^p])^{q/p} & \text{for } 1 \leq p < \infty, q \leq 1 \\ (\text{ess sup } |A - B|)^q & \text{for } p = \infty, q \leq 1 \end{cases}$$

from the metrics $M^{(p)}$ in (23).

2. $D \implies D/(1 + D)$. This is a standard way of creating a bounded p.q.-metric.
3. $D_1, D_2 \implies \max\{D_1, D_2\}$ or $D_1, D_2 \implies D_1 + D_2$. This transformations can be used to symmetrize p.q.-metrics: $D(A, B) + D(B, A)$ or $\max\{D(A, B), D(B, A)\}$.
4. A generalization of the previous: $\{D_v : v \in Y\} \implies \sup\{D_v\}$ and $\{D_v : v \in Y\} \implies E[D_V]$, where $\{D_v : v \in Y\}$ is a family of p.q.-metrics, and V designates a random entity distributed as (Y, Σ_Y, m) , so that

$$D(A, B) = \int_{v \in Y} D_v(A, B) dm(v).$$

We have discussed in Section 5.2.2 how such a procedure leads from our “classification” p.q.-metrics D_C to “separation” p.q.-metrics D_S .

6. NON-DISTANCE TESTS

The general principle of constructing tests for selective influences presented in Section 3.4.1 does not only lead to distance-type tests. In this section we will consider two examples, one

proposed previously and one new, of tests in which the functionals $g(H_\Phi)$ mentioned in Section 3.4.1 are, respectively, two-argument but not distance-type, and multiple-argument ones. Recall that the tests in question are only necessary conditions for selective influences (in the form of the diagram 8).

6.1. Cosphericity test

Given a hypothetical JDC-vector

$$H = \left(H_{x_1}^{\alpha_1}, \dots, H_{x_{k_1}}^{\alpha_{k_1}}, \dots, H_{x_1}^{\alpha_n}, \dots, H_{x_{k_n}}^{\alpha_n} \right)$$

with real-valued random variables, the following statement Θ should be satisfied: for any quadruple of factor points $\{x^\alpha, y^\beta, u^\alpha, v^\beta\}$ with $\alpha \neq \beta$ such that for some treatments $\phi_1, \phi_2, \phi_3, \phi_4 \in T$,

$$\{x^\alpha, y^\beta\} \subset \phi_1, \{x^\alpha, v^\beta\} \subset \phi_2, \{u^\alpha, y^\beta\} \subset \phi_3, \{u^\alpha, v^\beta\} \subset \phi_4,$$

we have

$$\begin{aligned} & \left| \rho_{x^\alpha y^\beta} \rho_{x^\alpha v^\beta} - \rho_{u^\alpha y^\beta} \rho_{u^\alpha v^\beta} \right| \\ & \leq \sqrt{1 - \rho_{x^\alpha y^\beta}^2} \sqrt{1 - \rho_{x^\alpha v^\beta}^2} + \sqrt{1 - \rho_{u^\alpha y^\beta}^2} \sqrt{1 - \rho_{u^\alpha v^\beta}^2}, \end{aligned}$$

where $\rho_{x^\alpha y^\beta}$ denotes the correlation between H_{x^α} and H_{y^β} , $\rho_{x^\alpha u^\beta}$ denotes the correlation between H_{x^α} and H_{u^β} , etc. Ergo, if the inequality is violated for at least one such a quadruple of factor points, the JDC-vector cannot exist, and the diagram of selective influences 8 should be rejected. For numerous illustrations see Kujala and Dzhamfarov (2008), where this test has been proposed, and where it is also shown that for two bivariate normally distributed variables in a 2×2 factorial design this test is both a necessary and sufficient condition for selective influences.

6.2. Diversity Test

The p.q.-metric $P^{(2)}$ introduced in Section 5 lends itself to an interesting generalization. Let \mathcal{R} be a set of jointly distributed random variables, each having $\{1, 2, \dots, s\}$ as its set of possible values. Define

$$\begin{aligned} P^{(s)} & \left[\left(R_1^1, \dots, R_1^{k_1} \right) \dots \left(R_i^1, \dots, R_i^{k_i} \right) \dots \left(R_s^1, \dots, R_s^{k_s} \right) \right] \\ & = \Pr \left[R_i^j = i, \text{ for } j = 1, \dots, k_i \text{ and } i = 1, \dots, s \right]. \end{aligned}$$

In particular,

$$\Pr \left[R_1 = 1, \dots, R_s = s \right] = P^{(s)} \left[(R_1) \dots (R_s) \right].$$

It is easy to show that the latter is a generalized p.q.-distance, in the sense of satisfying the following two properties: for any $R_1, \dots, R_s, R \in \mathcal{R}$,

1. (generalized premetric) $P^{(s)} \left[(R_1) \dots (R_s) \right]$ is nonnegative, and it is zero if any two of R_1, \dots, R_s are identical.

2. (simplicial inequality):

$$P^{(s)} \left[(R_1) \dots (R_s) \right] \leq \sum_{i=1}^s P^{(s)} \left[(R_1) \dots (R) \dots (R_s) \right],$$

where in the i th summand on the right, R_i in the sequence $(R_1) \dots (R_i) \dots (R_s)$ is replaced with R ($i = 1, \dots, s$), the rest of the sequence remaining intact.¹⁶

The generalized premetric property is obvious. To avoid cumbersome notation, let us prove the simplicial inequality for $s = 3$, the generalization to arbitrary s being straightforward. We drop in $P^{(3)}$ the parentheses around singletons: $P^{(3)} [R_1 R_2 R_3]$, $P^{(3)} [R_1 (R_2, R) R_3]$, etc. The simplicial inequality in question is

$$P^{(3)} [R_1 R_2 R_3] \leq P^{(3)} [R R_2 R_3] + P^{(3)} [R_1 R R_3] + P^{(3)} [R_1 R_2 R].$$

We have

$$\begin{aligned} & P^{(3)} [R_1 R_2 R_3] \\ & = P^{(3)} [(R_1, R) R_2 R_3] + P^{(3)} [R_1 (R_2, R) R_3] + P^{(3)} [R_1 R_2 (R_3, R)], \end{aligned}$$

$$\begin{aligned} & P^{(3)} [R R_2 R_3] \\ & = P^{(3)} [(R_1, R) R_2 R_3] + P^{(3)} [R (R_1, R_2) R_3] + P^{(3)} [R R_2 (R_1, R_3)], \end{aligned}$$

and analogously for $P^{(3)} [R_1 R R_3]$ and $P^{(3)} [R_1 R_2 R]$. Then

$$\begin{aligned} & P^{(3)} [R R_2 R_3] + P^{(3)} [R_1 R R_3] + P^{(3)} [R_1 R_2 R] - P^{(3)} [R_1 R_2 R_3] \\ & = P^{(3)} [R (R_1, R_2) R_3] + P^{(3)} [R R_2 (R_1, R_3)] \\ & \quad + P^{(3)} [(R_1, R_2) R R_3] + P^{(3)} [R_1 R (R_2, R_3)] \\ & \quad + P^{(3)} [(R_1, R_3) R_2 R] + P^{(3)} [R_1 (R_2, R_3) R] \geq 0. \end{aligned}$$

We call $P^{(s)}$ a *diversity* function. To use this function for a test of selective influences, for each random variable H_{x^α} in the hypothetical JDC-vector H we partition the set of its possible values \mathcal{A}_{x^α} into s pairwise disjoint subsets $\mathcal{A}_{x^\alpha}^1, \dots, \mathcal{A}_{x^\alpha}^s$, and we transform H_{x^α} as

$$R_{x^\alpha} = \begin{cases} 1 & \text{if } H_{x^\alpha} \in \mathcal{A}_{x^\alpha}^1, \\ \vdots & \vdots \\ s & \text{if } H_{x^\alpha} \in \mathcal{A}_{x^\alpha}^s. \end{cases}$$

Define

$$Dx_1^{\mu_1} \dots x_s^{\mu_s} = P^{(s)} \left[R_{x_1^{\mu_1}} \dots R_{x_s^{\mu_s}} \right].$$

¹⁶ With the addition of permutation-invariance, functions $\mathcal{R}^s \rightarrow \mathbb{R}$ (with \mathcal{R} an arbitrary set) satisfying these properties are sometimes called $(s-1)$ -semimetrics (Deza & Rosenberg, 2000); with the addition of the property that $P^{(s)} > 0$ if no two arguments thereof are equal, they become $(s-1)$ -metrics.

Let us restrict the consideration to $s = 3$ again. Assuming all factor points mentioned below belong to $\bigcup \Phi$, and given a triadic chain of factor points $t = x^\alpha y^\beta z^\gamma$ (with the elements pairwise distinct), we define a certain set of triadic chains referred to as a *polyhedral set* over t .

1. For any triadic chain $t = x^\alpha y^\beta z^\gamma$ ($x^\alpha \neq y^\beta \neq z^\gamma \neq x^\alpha$) and any $u^\mu \notin \{x^\alpha, y^\beta, z^\gamma\}$, the set $\{u^\mu y^\beta z^\gamma, x^\alpha u^\mu z^\gamma, x^\alpha y^\beta u^\mu\}$ is a polyhedral set over t ;
2. For any triadic chains t and t' , if \mathfrak{P} is a polyhedral set over t , and \mathfrak{P}' is a polyhedral set over any $t' \in \mathfrak{P}$, then the set $(\mathfrak{P} - \{t'\}) \cup \mathfrak{P}'$ is a polyhedral set over t .
3. Any polyhedral set over any triadic chain t is obtained by a finite number of applications of 1 and 2 above.

We call such a set polyhedral because if one interprets each element of it as a list of vertices forming a (triangular) face, then the whole set, combined with the root face t , forms a complete polyhedron.

A polyhedral set \mathfrak{P} over $t = x^\alpha y^\beta z^\gamma$ is called *treatment-realizable* if each element (triadic chain) that belongs to $\mathfrak{P} \cup \{t\}$ consists of elements of some treatment $\phi \in T$ (which implies, in particular, $\alpha \neq \beta \neq \gamma \neq \alpha$). The diversity test for selective influences consists in checking the compliance of the hypothetical JDC-vector with the following statement: for any treatment-realizable polyhedral set \mathfrak{P} over $x_1^{\mu_1} x_2^{\mu_2} x_3^{\mu_3}$,

$$Dx_1^{\mu_1} x_2^{\mu_2} x_3^{\mu_3} \leq \sum_{x_i^{\mu_i} x_j^{\mu_j} x_k^{\mu_k} \in \mathfrak{P}} Dx_i^{\mu_i} x_j^{\mu_j} x_k^{\mu_k}. \quad (24)$$

The inequality trivially follows from the simplicial inequality and the definition of \mathfrak{P} .

The classification p.q.-metric tests considered earlier form a special case of the diversity tests. For complete analogy one should replace chains in the formulation of the $P^{(2)}$ -based tests with a *polygonal set* \mathfrak{P} of pairs of factor points (*dipoles*) over a given dipole $d = x^\alpha y^\beta$ ($x^\alpha \neq y^\beta$). This set is defined as a set obtainable by repeated applications of the following two rules:

1. for any $d = x^\alpha y^\beta$ ($x^\alpha \neq y^\beta$) and any $u^\mu \notin \{x^\alpha, y^\beta\}$, the set $\{u^\mu y^\beta, x^\alpha u^\mu\}$ is a polygonal set over d ;
2. if \mathfrak{P} is a polygonal set over d , and \mathfrak{P}' is a polygonal set over any $d' \in \mathfrak{P}$, then the set $(\mathfrak{P} - \{d'\}) \cup \mathfrak{P}'$ is a polygonal set over d .

The generalization to $s > 3$ involves *polytopal sets* of s -element chains and is conceptually straightforward. The notion of an irreducible chain is also generalizable to polytopal sets, but we are not going to discuss this and related issues here: the diversity function and diversity-based tests form a rich topic that deserves a special investigation.

Example 6.1. Let $\alpha, \beta, \gamma, \delta$ be binary (1/2) factors, and let the set of allowable treatments T consist of all combinations of the factor points subject to the following constraint: $\{1^\alpha, 1^\beta, 2^\gamma, 1^\delta\}$ is the only treatment in T of the forms $\{1^\alpha, 1^\beta, 2^\gamma, v^\delta\}$, $\{1^\alpha, 1^\beta, v^\gamma, 1^\delta\}$, $\{1^\alpha, v^\beta, 2^\gamma, 1^\delta\}$, and $\{v^\alpha, 1^\beta, 2^\gamma, 1^\delta\}$. Let

the random variables A, B, C, D in the hypothetical diagram $(A, B, C, D) \leftarrow \rho (\alpha, \beta, \gamma, \delta)$ each have three values, denoted 1, 2, 3, and let the distributions of (A, B, C, D) be as shown in the tables, with all omitted joint probabilities being zero:

α	β	γ	δ	A	B	C	D	Pr
x	y	z	u	$:$	$:$	$:$	$:$	$:$
				1	2	3	1	1/3
				1	2	3	2	1/3
				1	2	3	3	1/3
				$:$	$:$	$:$	$:$	$:$

α	β	γ	δ	A	B	C	D	Pr
1	1	2	1	$:$	$:$	$:$	$:$	$:$
				1	2	3	1	1/2
				1	2	3	2	1/2
				1	2	3	3	0
				$:$	$:$	$:$	$:$	$:$

where $\{x^\alpha, y^\beta, z^\gamma, u^\gamma\}$ is any treatment in T other than $\{1^\alpha, 1^\beta, 2^\gamma, 1^\gamma\}$. It is easy to check that the 3-marginals (hence also all lower-order marginals) of the distributions satisfy marginal selectivity. One can also check that $\{1^\alpha 1^\beta 1^\delta, 1^\alpha 1^\gamma 1^\delta, 1^\beta 1^\gamma 1^\delta\}$ is a polyhedral set (in fact, the simplest one, forming a tetrahedron with vertices $1^\alpha, 1^\beta, 1^\gamma, 1^\delta$). This polyhedral set is treatment-realizable, because

$$\begin{aligned} \{1^\alpha, 1^\beta, 1^\gamma\} &\subset \{1^\alpha, 1^\beta, 1^\gamma, 2^\delta\}, & \{1^\alpha, 1^\beta, 1^\delta\} &\subset \{1^\alpha, 1^\beta, 2^\gamma, 1^\delta\}, \\ \{1^\alpha, 1^\gamma, 1^\delta\} &\subset \{1^\alpha, 2^\beta, 1^\gamma, 1^\delta\}, & \{1^\beta, 1^\gamma, 1^\delta\} &\subset \{2^\alpha, 1^\beta, 1^\gamma, 1^\delta\}. \end{aligned}$$

Putting

$$\begin{aligned} D1^\alpha 1^\beta 1^\gamma &= P^{(3)} [H_{1^\alpha} H_{1^\beta} H_{1^\gamma}] \\ &= \Pr [\{A = 1, B = 2, C = 3\} (1^\alpha, 1^\beta, 1^\gamma, 2^\delta)] = 1, \end{aligned}$$

$$\begin{aligned} D1^\alpha 1^\beta 1^\delta &= P^{(3)} [H_{1^\alpha} H_{1^\beta} H_{1^\delta}] \\ &= \Pr [\{A = 1, B = 2, D = 3\} (1^\alpha, 1^\beta, 2^\gamma, 1^\delta)] = 0, \end{aligned}$$

$$\begin{aligned} D1^\alpha 1^\delta 1^\gamma &= P^{(3)} [H_{1^\alpha} H_{1^\delta} H_{1^\gamma}] \\ &= \Pr [\{A = 1, D = 2, C = 3\} (1^\alpha, 2^\beta, 1^\gamma, 1^\delta)] = \frac{1}{3}, \end{aligned}$$

$$\begin{aligned} D1^\delta 1^\beta 1^\gamma &= P^{(3)} [H_{1^\delta} H_{1^\beta} H_{1^\gamma}] \\ &= \Pr [\{D = 1, B = 2, C = 3\} (2^\alpha, 1^\beta, 1^\gamma, 1^\delta)] = \frac{1}{3}, \end{aligned}$$

where H_{x^μ} are elements of the hypothetical JDC-vector, we see that the simplicial inequality is violated:

$$1 = D1^\alpha 1^\beta 1^\gamma > D1^\alpha 1^\beta 1^\delta + D1^\alpha 1^\delta 1^\gamma + D1^\delta 1^\beta 1^\gamma = \frac{2}{3}.$$

This rules out the possibility of $(A, B, C, D) \leftarrow \rho (\alpha, \beta, \gamma, \delta)$. \square

7. CONCLUSION

Selectiveness in the influences exerted by a set of inputs upon a set of random and stochastically interdependent outputs is a critical feature of many psychological models, often built into the very language of these models. We speak of an internal representation of a given stimulus, as separate from an internal representation of another stimulus, even if these representations are

considered random entities and they are not independent. We speak of decompositions of response time into signal-dependent and signal-independent components, or into a perceptual stage (influenced by stimuli) and a memory-search stage (influenced by the number of memorized items), without necessarily assuming that the two components or stages are stochastically independent. Moreover, the same as with theory of measurement and model selection studies, the issue of selective probabilistic influences, while born within psychology and motivated by psychological theorizing, pertains in fact to any area of empirical science dealing with inputs and random outputs.

In this paper, we have described the fundamental Joint Distribution Criterion for selective influences, and proposed a direct application of this criterion to random variables with finite numbers of values, the Linear Feasibility Test for selective influences. This test can be performed by means of standard linear programming. Due to the fact that any random output can be discretized, the Linear Feasibility Test is universally applicable, although one should keep in mind that if a diagram of selective influences is upheld by the test at some discretization, it may be rejected at a finer or non-nested discretization (but not at a coarser one).

Based on the Joint Distribution Criterion we have also formulated a general scheme for constructing various necessary conditions (tests) for selective influences. Among the tests thus generated is a wide spectrum of distance-type tests and some other tests described in the paper. The results of some of these tests (e.g., all those involving expected values) are not invariant with respect to factor-point-specific transformations of the random outputs, which allows one to expand each of such tests into an infinity of different tests for different transformations.

The abundance of different tests which we now have at our disposal poses new problems. The Linear Feasibility Test is superior to other tests as it allows one to prove (rather than only

disprove) the adherence of a system of inputs and outputs to a given diagram of selective influences (for a given discretization, if one is involved). It is possible, however, that discretization is not desirable, or the size of the problem is too large to be handled by available computational methods. In these cases one faces the problem of devising an optimal, or at least systematic way of applying a sequence of different necessary conditions, such as distance-type tests. Let us call a test T_1 stronger than test T_2 with respect to a given diagram of selective influences if the latter cannot be upheld by T_1 and rejected by T_2 , while the reverse is possible. Thus, in Kujala and Dzhamalov (2008) it is shown that the cosphericity test (Section 6.1) is stronger than the Minkowski distance test with $p = 2$ (Section 5.2.1). We know very little, however, about the comparative strengths of different tests on a broader scale.

The problem of devising optimal strategies of sequential testing arises also within the confines of a particular class of tests. Thus, the classification test (Sections 5.1 and 5.2.2) and the diversity test (Section 6.2) can be used repeatedly, each time with a different choice of the partitions of the random outputs' domains. We do not know at present how to organize the sequences of these choices optimally. In the case of the Minkowski distance test we do not know in which order one should use different values of p and different factor-point-specific transformations of the random variables. The latter also applies to the nonlinear transformations in the cosphericity test.

Finally, adaptation of the population-level tests to data analysis is another problem to be addressed by future research. Although sample-level procedures corresponding to our tests seem conceptually straightforward (Section 3.4.2), the issues of statistical power and statistical interdependence compound the problems of comparative strength of the tests and optimal strategy of sequential testing.

Appendix A: GENERALIZATIONS TO ARBITRARY SETS

Random Entities and Variables

For the purposes of this paper it is convenient to view a *random entity* A as a quadruple $(A, \mathcal{A}, \Sigma, \mu)$, where A is a unique name, \mathcal{A} is a nonempty set (of values of A), Σ is a sigma algebra of subsets of \mathcal{A} (called *measurable* subsets), and μ is a probability measure on Σ with the interpretation that $\mu(a)$ for any $a \in \Sigma$ is the probability with which A falls within $a \in \mathcal{A}$. (\mathcal{A}, Σ) is referred to as the *observation space* for A . We call the probability space $(\mathcal{A}, \Sigma, \mu)$ the *distribution* for A and say that A is *distributed as* $(\mathcal{A}, \Sigma, \mu)$. The inclusion of the label A is needed to ensure an unlimited collection of distinct random entities with the same distribution. If two random entities A and A' have the same distribution, we write $A \sim A'$. If A and B are distributed as, respectively, $(\mathcal{A}, \Sigma_A, \mu)$ and $(\mathcal{B}, \Sigma_B, \nu)$, then we say $B \sim f(A)$ if $f : \mathcal{A} \rightarrow \mathcal{B}$ is such that $b \in \Sigma_B$ implies $f^{-1}(b) \in \Sigma_A$ and $\nu(b) = \mu(f^{-1}(b))$, ν being referred to as the *induced measure* (with respect to μ, f), and the function f being said to be $(\mathcal{A}, \Sigma_A, \mu) - (\mathcal{B}, \Sigma_B, \nu)$ -*measurable*.

With any indexed set of random entities $\{A_\omega\}_{\omega \in \Omega}$ each of which is distributed as $(\mathcal{A}_\omega, \Sigma_\omega, \mu_\omega)$, $\omega \in \Omega$, we associate its

“natural” observation space (\mathcal{A}, Σ) , with $\mathcal{A} = \prod_{\omega \in \Omega} \mathcal{A}_\omega$ (Cartesian product) and $\Sigma = \bigotimes_{\omega \in \Omega} \Sigma_\omega$ being the smallest sigma algebra containing all sets of the form $a_\omega \times \prod_{i \in \Omega - \{\omega\}} \mathcal{A}_i$, $a_\omega \in \Sigma_\omega$. We say that the random entities in $\{A_\omega\}_{\omega \in \Omega}$ *possess a joint distribution* if $\{A_\omega\}_{\omega \in \Omega}$ is a random entity distributed as $(\mathcal{A}, \Sigma, \mu)$ with $\mu(a_\omega \times \prod_{i \in \Omega - \{\omega\}} \mathcal{A}_i) = \mu_\omega(a_\omega)$. Every subset $\Omega' \subset \Omega$ possesses a *marginal distribution* $(\prod_{\omega \in \Omega'} \mathcal{A}_\omega, \bigotimes_{\omega \in \Omega'} \Sigma_\omega, \mu')$, where $\mu'(a) = \mu(a \times \prod_{i \in \Omega - \Omega'} \mathcal{A}_i)$, for all $a \in \bigotimes_{\omega \in \Omega'} \Sigma_\omega$.¹⁷

¹⁷ The standard definition of a random entity (also called “random element” or simply “random variable”) is a measurable function from a sample space to an observation space. The present terminology can be reconciled with this view by considering $(\{A\} \times \mathcal{A}, \{\{A\} \times a : a \in \Sigma\}, \nu)$ a sample space, (\mathcal{A}, Σ) an observation space, and A the projection function $\{A\} \times \mathcal{A} \rightarrow \mathcal{A}$. In the case of jointly distributed random entities, $A = \{A_\omega\}_{\omega \in \Omega}$, each of them, with an observation space $(\mathcal{A}_\omega, \Sigma_\omega)$, can be defined as the projection function $\{A\} \times \mathcal{A} \rightarrow \mathcal{A}_\omega$. We do not, however, assume a common sample space for all random entities being considered. The notion of a sample space is a source of conceptual confusions, the chief one being the notion that there is only one sample space “in this universe,” so that any set of random entities possesses a

Remark A.1. Note that the elements of the Cartesian product $\prod_{\omega \in \Omega} \mathcal{A}_\omega$ are *choice functions* $\Omega \rightarrow \bigcup_{\omega \in \Omega} \mathcal{A}_\omega$, that is, they are sets of pairs of the form (ω, a) , $\omega \in \Omega$, $a \in \mathcal{A}_\omega$. This means that the indexation of $\{A_\omega\}_{\omega \in \Omega}$ is part of the identity of $\mathcal{A} = \prod_{\omega \in \Omega} \mathcal{A}_\omega$, hence also of the distribution of $A = \{A_\omega\}_{\omega \in \Omega}$. Ideally, only the “ordinal structure” of the indexing set Ω should matter, and this can be ensured by agreeing that Ω is always an initial segment of the class of ordinal numbers. With these conventions in mind, $\{A_\omega\}_{\omega \in \Omega}$ can be viewed as generalizing the notion of a finite vector (although it is convenient not to complicate notation to reflect this fact). For sets of jointly distributed and identically indexed random entities, the relation $\{A_\omega\}_{\omega \in \Omega} \sim \{B_\omega\}_{\omega \in \Omega}$ should always be understood in the sense of “corresponding indices,” implying, in particular, $\{A_\omega\}_{\omega \in \Omega'} \sim \{B_\omega\}_{\omega \in \Omega'}$ for any subset Ω' of Ω .

The equality $A_1 = A_2$ in the present context means that the two random entities have a common observation space (\mathcal{A}, Σ) , and that $\{A_1, A_2\}$ is a jointly distributed random entity with measure μ such that $\mu(\{(a_1, a_2) \in \mathcal{A} \times \mathcal{A} : a_1 = a_2\}) = 1$ (this corresponds to the equality “almost surely” in the traditional terminology). We also follow the common practice of using equality to replace “is” or “denotes” in definitions and abbreviations, such as $A = \{A_\omega\}_{\omega \in \Omega}$. The two meanings of equality are easily distinguished by context.

A *random variable* is a special case of random entity. Its definition can be given as follows: (i) if \mathcal{A} is countable, Σ is the power set of \mathcal{A} , then a random entity distributed as $(\mathcal{A}, \Sigma, \mu)$ is a random variable; (ii) if \mathcal{A} is an interval of reals, Σ is the Lebesgue sigma-algebra on \mathcal{A} , then a random entity distributed as $(\mathcal{A}, \Sigma, \mu)$ is a random variable; (iii) any jointly distributed vector (A_1, \dots, A_n) with all components random variables is a random variable. The notion thus defined is more general than in the main text, but the theory presented there applies with no modifications.

Lemma A.2. A set $\{A_\omega\}_{\omega \in \Omega}$ of random entities possesses a joint distribution if and only if there is a random entity R distributed as a probability space $(\mathcal{R}, \Sigma_{\mathcal{R}}, \nu)$ and some functions $\{f_\omega : \mathcal{R} \rightarrow \mathcal{A}_\omega\}_{\omega \in \Omega}$, such that $\{A_\omega\}_{\omega \in \Omega} = \{f_\omega(R)\}_{\omega \in \Omega}$.

Proof. (Note that the formulation implies that all the functions involved are appropriately measurable.) To show sufficiency, observe that the induced measure μ of any set of the form $\prod_{\omega \in N} \mathcal{A}_\omega \times \prod_{\omega \in \Omega - N} \mathcal{A}_\omega$, where N is a finite subset of Ω and $\mathbf{a}_\omega \in \Sigma_\omega$ for $\omega \in N$, is $\nu(\bigcap_{\omega \in N} f_\omega^{-1}(\mathbf{a}_\omega))$, and this measure is uniquely extended to $\bigotimes_{\omega \in \Omega} \Sigma_\omega$. To show necessity, put $R = \{A_\omega : \omega \in \Omega\}$ and, for every $\omega \in \Omega$, define $f_\omega : \mathcal{R} \rightarrow \mathcal{A}_\omega$ to be the (obviously measurable) projection $f_\omega : \prod_{\omega \in \Omega} \mathcal{A}_\omega \rightarrow \mathcal{A}_\omega$. \square

Corollary A.3. If Ω is finite and $\{A_\omega\}_{\omega \in \Omega}$ is a set of random variables, then R in Lemma A.2 can be chosen to be a random variable. Moreover, R can be chosen arbitrarily, as any continuously (atomlessly) distributed random variable (e.g., uniformly distributed between 0 and 1).

Proof. The first statement follows from the fact that $R = \{A_\omega\}_{\omega \in \Omega}$ in the necessity part of Lemma A.2 is then a random variable. The second statement follows from Theorem 1 in Dzhafarov & Gluhovsky, 2006, based on a general result for standard Borel spaces (e.g., in Kechris, 1995, p. 116). \square

Selective influences and JDC

A *factor* is defined as a nonempty set of *factor points* with a unique name: the notation used is $x^\alpha = \{x, \alpha\}$. Let Φ be a nonempty set of factors, and let $T \subset \prod \Phi$ be a nonempty set of *treatments*. Note that any treatment $\phi \in T$ is a function $\phi : \Phi \rightarrow \bigcup \Phi$, so $\phi(\alpha)$ denotes the factor point x^α of the factor α which belongs to the treatment ϕ . (The notation for $\phi(\alpha)$ used in the main text is $\phi_{\{\alpha\}}$.)

Let Ω be an indexing set for a set of random entities $\{R_\omega\}_{\omega \in \Omega}$. A diagram of selective influences is a mapping $M : \Omega \rightarrow 2^\Phi$. For any such a diagram one can redefine the set of factors and the set of treatments in the following way. For every $\omega \in \Omega$, put

$$\omega^* = \{s^{\omega^*} : s \in \prod M(\omega)\},$$

if $M(\omega)$ is nonempty; if it is empty, put $\omega^* = \{\emptyset^{\omega^*}\}$. This establishes the bijective mapping $M^* : \Omega \rightarrow 2^{\Phi^*}$, where $\Phi^* = \{\omega^*\}_{\omega \in \Omega}$. For each treatment $\phi \in T$ we define the corresponding treatment ϕ^* as $\{s^{\omega^*} : s \in \phi \wedge s \in \prod M^*(\omega), \omega \in \Omega\}$. The set of all such treatments ϕ^* is denoted T^* . (In the main text the procedure just described is called *canonical rearrangement*.) In the following we omit asterisks and simply put $\Phi = \Omega$, replacing $M : \Omega \rightarrow 2^\Phi$ with the identity map $M : \Omega \rightarrow \Phi$.

Among several equivalent definitions of selective influences we choose here the one most immediately prompting the Joint Distribution Criterion (JDC).

Definition A.4. Let $\mathbb{A} = \{A_\phi\}_{\phi \in T}$, and $A_\phi = \{A_{\phi, \alpha}\}_{\alpha \in \Phi}$ for every $\phi \in T$. Let T be a set of treatments associated with a set of factors Φ . Let $A_{\phi, \alpha}$ for each α, ϕ be distributed as $(\mathcal{A}_{\phi(\alpha)}, \Sigma_{\phi(\alpha)}, \mu_{\phi, \alpha})$. We say that each $A_{\phi, \alpha}$ is *selectively influenced by α* ($\alpha \in \Phi, \phi \in T$), and write schematically $\mathbb{A} \leftarrow \Phi$, if there is a random entity R distributed as $(\mathcal{R}, \Sigma_{\mathcal{R}}, \nu)$ and some functions $\{f_{\alpha} : \mathcal{R} \rightarrow \mathcal{A}_{\alpha}\}_{\alpha \in \bigcup \Phi}$ such that $A_\phi = \{A_{\phi, \alpha}\}_{\alpha \in \Phi} \sim \{f_{\phi(\alpha)}(R)\}_{\alpha \in \Phi}$, for all $\phi \in T$.

Remark A.5. Note that the formulation implies that all the functions involved are appropriately measurable. Also, in $\{f_{\alpha} : \mathcal{R} \rightarrow \mathcal{A}_{\alpha}\}_{\alpha \in \bigcup \Phi}$ the set $\bigcup \Phi$ can be replaced with $\bigcup_{\phi \in T, \alpha \in \Phi} \phi(\alpha)$ if the latter is a proper subset of $\bigcup \Phi$ (and the same applies to the definition of H in the theorem below). We assume, however, that factor points never used in treatments can simply be deleted from the factors.

Remark A.6. In the main text we assume that $(\mathcal{A}_{\phi(\alpha)}, \Sigma_{\phi(\alpha)}) = (\mathcal{A}_\alpha, \Sigma_\alpha)$, that is, the observation space $(\mathcal{A}_\alpha, \Sigma_\alpha)$ of the entity $A_{\phi, \alpha}$ is the same across different treatments $\phi \in T$. In footnote 6 we mention that this constraint is not essential, as the random entities $A_{\phi, \alpha}$ can always be redefined to force $(\mathcal{A}_{\phi(\alpha)}, \Sigma_{\phi(\alpha)}) =$

$(\mathcal{A}_\alpha, \Sigma_\alpha)$ without affecting selective influence. This redefinition can be done in a variety of ways, the simplest one being to put

$$\mathcal{A}_\alpha = \bigcup_{\phi \in T} \{\phi(\alpha)\} \times \mathcal{A}_{\phi(\alpha)},$$

and let Σ_α be the smallest sigma-algebra containing $\{\{\phi(\alpha)\} \times \alpha : \alpha \in \Sigma_{\phi(\alpha)}, \phi \in T\}$. Define $g_{\phi(\alpha)} : \mathcal{A}_{\phi(\alpha)} \rightarrow \mathcal{A}_\alpha$ by $g_{\phi(\alpha)}(a) = (\phi(\alpha), a)$, for $a \in \mathcal{A}_{\phi(\alpha)}, \phi \in T, \alpha \in \Phi$. Then $A_{\phi, \alpha}^* = g_{\phi(\alpha)}(A_{\phi, \alpha})$ and $A_\Phi^* = \{A_{\phi, \alpha}^*\}_{\alpha \in \Phi}$ are the redefined random entities sought. Note that if $\mathbb{A} \dashv \vdash \Phi$, then $\mathbb{A}^* = \{A_\phi^*\}_{\phi \in T} \dashv \vdash \Phi$, because Definition A.4 applies to \mathbb{A}^* with the same R and with the composite functions $g_{x^\alpha} \circ f_{x^\alpha}$ replacing f_{x^α} , for all $x^\alpha \in \bigcup \Phi$. (In the terminology of the main text, g_{x^α} are factor-point-specific transformations.)

Theorem A.7 (JDC). *A necessary and sufficient condition for $\mathbb{A} \dashv \vdash \Phi$ in Definition A.4 is the existence of a set of jointly distributed random entities*

$$H = \{H_{x^\alpha}\}_{x^\alpha \in \bigcup \Phi}$$

(one random entity for each factor point of each factor), such that

$$\{H_{x^\alpha}\}_{x^\alpha \in \phi} \sim A_\phi$$

for every treatment $\phi \in T$.

Proof. Immediately follows from the definition and Lemma A.2. \square

Theorem A.8. *If $\bigcup \Phi$ in Definition A.4 is a finite set and $A_{\phi(\alpha)}$ is a random variable for every α, ϕ , then R can always be chosen to be a random variable. Moreover, R can be chosen arbitrarily, as any continuously (atomlessly) distributed random variable.*

Proof. Immediately follows from JDC and Corollary A.3. \square

Remark A.9. In Dzhaferov and Gluhovsky (2006) this inference was not made because JDC at that time was not explicitly formulated (outside quantum mechanics, see footnotes 11 and 13).

The three basic properties of selective influences listed in Section 3.3 trivially generalize to arbitrary sets of factors and random entities.

Distance-type tests

The principles of test construction (Section 3.4) and the logic of the distance-type tests in particular, apply without changes to arbitrary sets of factors. As to the random entities, some of the test measures are confined to discrete and/or real-valued variables (e.g., information-based and Minkowski-type ones), others (such as classification measures) are completely general.

We will use the notation and terminology adopted in Dzhaferov and Kujala (2010). Chains of factor points can be denoted by capital Roman letters, $X = x_1^{\alpha_1} \dots x_l^{\alpha_l}$. A subsequence of points belonging to a chain forms its *subchain*. A *concatenation* of two chains X and Y is written as XY . So, we can have chains

$x^\alpha X y^\beta, x^\alpha X Y y^\beta$, etc. The number of points in a chain X is its *cardinality*, $|X|$. For any treatment-realizable chain $X = x_1^{\alpha_1} \dots x_l^{\alpha_l}$, we write

$$DX = \sum_{i=1}^{l-1} Dx^{\alpha_i} x^{\alpha_{i+1}}$$

(with the understanding that the sum is zero if l is 0 or 1).

A treatment-realizable chain $u^\mu X v^\nu$ is called *compliant* (with the chain inequality) if $Du^\mu v^\nu \leq Du^\mu X v^\nu = Dx^{\alpha_1} x^{\alpha_l} + DX + Dx^{\alpha_l} x^{\alpha_1}$; it is called *contravening* (the chain inequality) if $Du^\mu v^\nu > Du^\mu X v^\nu$. The proofs of the two lemmas below are very similar, but it is convenient to keep them separate.

Lemma A.10. *If a treatment-realizable chain $X_0 = x_1^{\alpha_1} \dots x_l^{\alpha_l}$ ($l \geq 3$) is contravening, then it contains a contravening subchain in which no factor point occurs more than once.*

Proof. If $l = 3$ then the chain contains no factor point more than once, because otherwise it is not treatment-realizable. If $l > 3$, and X_0 contains factor points $x_i^{\alpha_i} = x_j^{\alpha_j}$, then it can be presented as $X_0 = x_1^{\alpha_1} \dots x_i^{\alpha_i} U x_j^{\alpha_j} \dots x_l^{\alpha_l}$, where U is some nonempty subchain (i may coincide with 1 or j coincide with l , but not both). But then $X_1 = x_1^{\alpha_1} \dots x_i^{\alpha_i} \dots x_l^{\alpha_l}$ is also treatment-realizable and contravening, because

$$\begin{aligned} Dx_1^{\alpha_1} x_l^{\alpha_l} &> DX_0 = Dx_1^{\alpha_1} \dots x_i^{\alpha_i} U x_j^{\alpha_j} \dots x_l^{\alpha_l} \\ &> Dx_1^{\alpha_1} \dots x_i^{\alpha_i} \dots x_l^{\alpha_l} = DX_1. \end{aligned}$$

If X_1 contains two equal factor points, then $3 \leq |X_1| < |X_0|$, and we can repeat the same procedure to obtain X_2 , etc. As the procedure has to stop at some X_l , this subchain will contain no factor point twice. \square

Lemma A.11. *If a treatment-realizable chain $X_0 = x_1^{\alpha_1} \dots x_l^{\alpha_l}$ ($l \geq 3$) is contravening, then it contains a contravening irreducible subchain.*

Proof. By the previous lemma, we can assume that every factor point in X_0 occurs no more than once. If $l = 3$, the chain X_0 itself is irreducible, because otherwise there would exist a treatment $\phi \in T$ that includes the elements of the chain, and this would make the chain compliant. If $l > 3$, and the chain X_0 is not irreducible, then it must contain a subchain $x_i^{\alpha_i} x_j^{\alpha_j}$ such that $j > i + 1$ and $\{x_i^{\alpha_i}, x_j^{\alpha_j}\}$ is part of some treatment $\phi \in T$. The chain then can be presented as $X_0 = x_1^{\alpha_1} \dots x_i^{\alpha_i} U x_j^{\alpha_j} \dots x_l^{\alpha_l}$, where U is some nonempty subchain (i may coincide with 1 or j with l , but not both). The subchain $x_i^{\alpha_i} U x_j^{\alpha_j}$ is clearly treatment-realizable. If it is contravening, then we replace X_0 with $X_1 = x_i^{\alpha_i} U x_j^{\alpha_j}$; if it is compliant, then we replace X_0 with $X_1 = x_1^{\alpha_1} \dots x_i^{\alpha_i} x_j^{\alpha_j} \dots x_l^{\alpha_l}$. In both cases we obtain a treatment-realizable subchain X_1 of X_0 such that $3 \leq |X_1| < |X_0|$, and X_1 is contravening: in the former case $X_1 = x_i^{\alpha_i} U x_j^{\alpha_j}$ is contravening by construction, in the latter case $Dx_i^{\alpha_i} U x_j^{\alpha_j} > Dx_i^{\alpha_i} x_j^{\alpha_j}$ whence

$$\begin{aligned} Dx_1^{\alpha_1} x_l^{\alpha_l} &> DX_0 = Dx_1^{\alpha_1} \dots x_i^{\alpha_i} U x_j^{\alpha_j} \dots x_l^{\alpha_l} \\ &> Dx_1^{\alpha_1} \dots x_i^{\alpha_i} x_j^{\alpha_j} \dots x_l^{\alpha_l} = DX_1. \end{aligned}$$

If X_1 is not irreducible, we can apply the same procedure to X_1 to obtain a contravening subchain X_2 with $3 \leq |X_2| < |X_1|$, and continue in this manner. Eventually we have to reach a contravening subchain X_t of X_0 such that $|X_t| \geq 3$ and the procedure cannot continue, indicating that X_t is irreducible. \square

REFERENCES

- Basoalto, R.M., & Percival, I.C. (2003). BellTest and CHSH experiments with more than two settings. *Journal of Physics A: Mathematical & General*, 36, 7411–7423.
- Bloxom, B. (1972). The simplex in pair comparisons. *Psychometrika*, 37, 119–136.
- Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. New York: Wiley.
- Deza, M.-M., & Rosenberg, I. G. (2000). n-Semimetrics. *European Journal of Combinatorics*, 21, 797–806.
- Dzhafarov, E.N. (1992). The structure of simple reaction time to step-function signals. *Journal of Mathematical Psychology*, 36, 235–268.
- Dzhafarov, E. N. (1997). Process representations and decompositions of response times. In A. A. J. Marley (Ed.), *Choice, decision and measurement: Essays in honor of R. Duncan Luce*, pp. 255–278. New York: Erlbaum.
- Dzhafarov, E.N. (1999). Conditionally selective dependence of random variables on external factors. *Journal of Mathematical Psychology*, 43, 123–157.
- Dzhafarov, E.N. (2001). Unconditionally selective dependence of random variables on external factors. *Journal of Mathematical Psychology*, 45, 421–451.
- Dzhafarov, E.N. (2002). Multidimensional Fechnerian scaling: Pairwise comparisons, regular minimality, and nonconstant self-similarity. *Journal of Mathematical Psychology*, 46, 583–608.
- Dzhafarov, E.N. (2003a). Selective influence through conditional independence. *Psychometrika*, 68, 7–26.
- Dzhafarov, E.N. (2003b). Thurstonian-type representations for “same-different” discriminations: Deterministic decisions and independent images. *Journal of Mathematical Psychology*, 47, 208–228.
- Dzhafarov, E.N. (2003c). Thurstonian-type representations for “same-different” discriminations: Probabilistic decisions and interdependent images. *Journal of Mathematical Psychology*, 47, 229–243.
- Dzhafarov, E.N. (2006) On the Law of Regular Minimality: Reply to Ennis. *Journal of Mathematical Psychology*, 50, 74–93.
- Dzhafarov, E.N. (2010). Dissimilarity, quasidistance, distance. *Journal of Mathematical Psychology*, 54, 334–337.
- Dzhafarov, E.N., & Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. *Journal of Mathematical Psychology*, 50, 390–401.
- Dzhafarov, E.N., & Kujala, J.V. (2010). The Joint Distribution Criterion and the Distance Tests for selective probabilistic causality. *Frontiers in Quantitative Psychology and Measurement*, 1:151 doi: 10.3389/fpsyg.2010.00151.
- Dzhafarov, E.N., Schweickert, R., & Sung, K. (2004). Mental architectures with selectively influenced but stochastically interdependent components. *Journal of Mathematical Psychology*, 48, 51–64.
- Fine, A. (1982a). Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics*, 23, 1306–1310.
- Fine, A. (1982b). Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48, 291–295.
- Kechris, A. S. (1995). *Classical Descriptive Set Theory*. New York: Springer.
- Kraskov, A., Stögbauer, H., Andrzejak, R.G., & Grassberger, P. (2003). Hierarchical Clustering Based on Mutual Information. *ArXiv q-bio/0311039*.
- Kujala, J. V., & Dzhafarov, E. N. (2008). Testing for selectivity in the dependence of random variables on external factors. *Journal of Mathematical Psychology*, 52, 128–144.
- Kujala, J.V., & Dzhafarov, E.N. (2009). Regular Minimality and Thurstonian-type modeling. *Journal of Mathematical Psychology*, 53, 486–501.
- Levine, M. V. (2003). Dimension in latent variable models. *Journal of Mathematical Psychology*, 47, 450–466.
- McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometrika Monographs*, 15.
- McDonald, R. P. (1982). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement*, 6, 379–396.
- Schweickert, R. (1982). The bias of an estimate of coupled slack in stochastic PERT networks. *Journal of Mathematical Psychology*, 26, 1–12.
- Schweickert, R., Fisher, D.L., & Goldstein, W.M. (2010). Additive factors and stages of mental processes in task networks. *Journal of Mathematical Psychology*, 54, 405–414.
- Schweickert, R., & Townsend, J. T. (1989). A trichotomy: Interactions of factors prolonging sequential and concurrent mental processes in stochastic discrete mental (PERT) networks. *Journal of Mathematical Psychology*, 33, 328–347.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders’ method. In W.G. Koster (Ed.), *Attention and Performance II. Acta Psychologica*, 30, 276–315.
- Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese*, 48, 191–199.
- Taylor, M.D. (1984). Separation metrics for real-valued random variables. *International Journal of Mathematics and Mathematical Sciences*, 7, 407–408.
- Taylor, M.D. (1985). New metrics for weak convergence of distribution functions. *Stochastica*, 9, 5–17.
- Thurstone, L. L. (1927b). A law of comparative judgments. *Psychological Review*, 34, 273–286.
- Townsend, J.T. & Ashby, G. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press.
- Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology*, 28, 363–400.
- Townsend, J.T., & Schweickert, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology*, 33, 309–327.
- Townsend, J. T., & Thomas, R. D. (1994). Stochastic dependencies in parallel and serial models: Effects on systems facto-

- rial interactions. *Journal of Mathematical Psychology*, 38, 1–34.
- Webster, R. (1994). *Convexity*. New York: Oxford University Press.
- Werner, R.F., & Wolf, M.M. (2001). All multipartite Bell correlation inequalities for two dichotomic observables per site. arXiv:quant-ph/0102024v1.
- Werner, R.F., & Wolf, M.M. (2001). Bell inequalities and entanglement. arXiv:quant-ph/0107093v2.